DOCUMENT RESUME

ED 283 870 TM 870 400

AUTHOR Davis, Wesley D.
TITLE An Empirical Assessment of Selected Software Purported to Raise SAT Scores Significantly When Utilized With Short-Term CAI on the Microcomputer.
PUB DATE Jul 85
NOTE 102p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Cognitive Style; College Bound Students; *College Entrance Examinations; *Computer Assisted Instruction; *Computer Software; Conditioning; Error of Measurement; Goodness of Fit; High Schools; Hypothesis Testing; Literature Reviews; Microcomputers; Pretests Posttests; Retention (Psychology); *Scores; Student Characteristics; *Test Coaching; Test Validity
IDENTIFIERS Chi Square Analysis; Florida (Escambia County); *Krell SAT Computer Software; Mann Whitney U Test; *Scholastic Aptitude Test

ABSTRACT

This study evaluated Krell's 1981-82 Scholastic Aptitude Test (SAT) preparatory series software purported to raise students' scores substantially after only a short term of computer-assisted instruction (CAI). Forty-eight college-bound juniors from Escambia County (Florida) were assigned to experimental and control groups. A two-phased pre- and post-test design was used. Both groups received 15 hours of CAI, but during different phases of the study. Data were collected through the administration of one SAT pretest, two SAT posttests, the Otis-Lennon School Ability Test, the Witkin Group Embedded Figures Test, and a student questionnaire for demographic information. Ten hypotheses concerning the effects of CAI, cognitive styles, socioeconomic status, gender, IQ, grade point average, parents' educational attainment, parents' occupational status, number of siblings, and college majors on SAT score gains were tested using non-parametric statistics. With one partial exception, all hypotheses were accepted in the null form. The results showed: (1) Krell SAT software utilized under controlled CAI conditions produces only modest group score gains, and (2) individual student score gains under these conditions, by contrast, may be high. That is, group SAT score gains will seldom exceed chance expectations and will be of little practical consequence. On the other hand, individual students may evidence significant gains, which is consistent with the publishers claim. The study includes a nine-page reference list. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED283870

AN EMPIRICAL ASSESSMENT OF SELECTED SOFTWARE PURPORTED TO RAISE SAT SCORES SIGNIFICANTLY WHEN UTILIZED WITH SHORT-TERM CAI ON THE MICROCOMPUTER

Wesley D. Davis, Ph.D.
Coordinator, Evaluation Services
Escambia County Public Schools
Pensacola, Florida

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

W. D. Davis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Copyright © July, 1985

TM 870 400

BEST COPY AVAILABLE

ABSTRACT

The purpose of this study was to assess empirically software from Krell's 1981/82 SAT preparatory series purported to raise students' scores substantially after only short-term CAI. The latter involved individualized student use of Apple IIe microcomputers in a controlled setting. These students were forty-eight college-bound juniors from Escambia County (Florida) who were assigned as Experimentals and Controls such that two stochastically equated groups reflecting matched pairs resulted.

A two-phased, pre- and post-test design was used. Phase I involved: 1) pre-testing with: a) the SAT, b) the Otis-Lennon School Ability Test, and c) the Witkin Group Embedded Figures Test; 2) fifteen hours of CAI with the Experimentals; and 3) post-testing with the SAT for outcome measures and baseline comparison scores. Phase II involved: 1) fifteen hours of CAI with the Controls, and 2) a second post-testing of both groups. The latter provided a 28-day retention measure for the Experimentals and Phase I replicated outcome scores for the Controls. Additionally, a student questionnaire was used to obtain demographic data in re: a) gender, b) number of siblings, c) educational level of parents, d) occupational status of parents, e) family income, f) G.P.A., and g) proposed college major.

Data analyses utilized nonparametric statistics in testing ten hypotheses. With one partial exception, all were accepted in the null form. On the whole, post-CAI group mean SAT score gains were modest and deemed to be of limited practical consequence. These findings are consistent with previous reports of SAT coaching efforts cited in the literature. Recommendations suggesting caution and scrutiny relative to publishers' claims of product effectiveness are offered. Further research is encouraged, particularly as related to individual student score gains and the development of a predictive profile.

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES viii

Chapter Page

I. INTRODUCTION TO THE STUDY 1

A. Statement of the Problem 1

B. Purpose of the Study 2

C. Significance of the Study 3

D. Conceptual Framework 4

E. Assumptions 7

F. Limitations 7

G. Hypotheses 8

II. REVIEW OF RELATED LITERATURE 10

A. Historical Use of College Admission Scores 10

B. Effects of Tutoring upon SAT Scores 12

C. Attacks upon the SAT 14

D. Defense of the SAT 15

E. Potential Impact of the Microcomputer 16

F. Computer-assisted Instruction 17

G. Cognitive Styles 20

H. Summary 21

III. METHODOLOGY 23

A. Population 23

B. Data Collection 25

1. Design 25

2. Equipment and Materials 27

3. Instruments 29

C. Data Analysis 32

IV. FINDINGS OF THE STUDY 36

A. Representativeness of the Sample 36

B. Application of the Design 41

C. Hypothesis Testing 42

D. Individual Score Changes 63

E. Retention 70

F. Correlational Relationships 72

G. Summary 72

V. SUMMARY, FINDINGS, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS 74

A. Summary 74

B. Findings 76

C. Conclusions 77

D. Discussion 78

E. Recommendation to the Field 83

F. Recommendation for Further Research 84

REFERENCES 85

LIST OF TABLES

Table		Page
1.	Chi-square Goodness-of-Fit Test Applied to SAT Total Pre-test Scores to Demonstrate that the Scores Are Normally Distributed and Representative of the Larger Population	38
2.	Mann-Whitney U Applied to SAT Total Pre-test Scores When Separated by Group	40
3.	Mann-Whitney U Applied to Otis-Lennon School Ability Test I.Q. Scores When Separated by Group	40
4.	Mann-Whitney U Applied to SAT Pre- and Post-test Score Changes Following CAI	43
5.	SAT Pre- and Post-test Group Mean Score Changes Reported as Gain or Loss	45
6.	Chi-square Applied to Differences Among SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated as Subjects Who Are Field-dependent (F-D) Versus Field-independent (F-I) Based Upon Their Scores From the GEFT	51
7.	Mann-Whitney U Applied to SAT Post-test Scores Reported as Gain or Loss Relative to Pre-test Scores When the Group Is Separated for Family Income Above or Below the Group Median	52
8.	Mann-Whitney U Applied to SAT Pre-test Score Differences When the Group Is Separated by Gender	54
9.	Mann-Whitney U Applied to SAT Post-test Score Differences When the Group Is Separated by Gender	56
10.	Chi-square Applied to Gender Differences Among SAT Post-test Scores When Separated Above and Below the Group Mean	58
11.	Mann-Whitney U Applied to SAT Post-test Scores Reported as Gain or Loss Relative to Pre-test Scores When the Group Is Separated for I.Q.s Above or Below the Group Mean	58
12.	Mann-Whitney U Applied to SAT Post-test Scores Reported as Gain or Loss Relative to Pre-test Scores When the Group Is Separated for G.P.A.s Above or	

Below the Group Mean 59

13. Chi-square Applied to SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated Relative to Educational Level of the Parents 60

14. Chi-square Applied to SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated Relative to Occupation of the Parents 62

15. Chi-square Applied to SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated Relative to Subjects' Number of Siblings 64

16. Chi-square Applied to SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated Relative to Subjects' Proposed College Major 64

17. Standard Error of Measurement Derived From SAT Pre- and Post-test Score Differences Following CAI 71

18. Mann-Whitney U Applied to Group 1 SAT Post-test Retention Scores One Month Following Completion of the CAI 71

19. Spearman Rho Applied to SAT Post-test Scores and O-LSAT (I.Q.), G.P.A., and Family Income (F.I.) to Determine the Degree of Correlational Relationship 73

LIST OF FIGURES

Figure — Page

1. SAT Verbal Pre- and Post-test Group Mean Score Changes 46

2. SAT Math Pre- and Post-test Group Mean Score Changes 48

3. SAT Total Pre- and Post-test Group Mean Score Changes 49

4. Number of Groups 1 and 2 Students Evidencing a Gain or Loss on Either SAT Verbal or Math When Pre- and Post-test Scores Are Compared Following CAI 65

5. Group 1 Students 28 and 29 Evidencing Extreme SAT Verbal Post-test Score Changes Following CAI 67

6. Group 2 Student 27 Evidencing Extreme SAT Math Post-test Score Change Following CAI 68

7. Group 2 Student 27 Evidencing Extreme SAT Total Post-test Score Change Following CAI 69

CHAPTER I

INTRODUCTION TO THE STUDY

Statement of the Problem

Attempts to raise Scholastic Aptitude Test (SAT) scores through a variety of coaching techniques have now been investigated for more than forty years (Messick & Jungeblut, 1981). The general conclusion has been that score increases, irrespective of the technique employed, seldom represent gains of more than one or two items correct over non-intervention expectations (Alderman & Powers, 1980). Consequently, members of the College Entrance Examination Board have for years taken the position that the minimal gains related to the time and expense to students and parents associated with coaching activities have not been worth the investment. Moreover, the Board has further claimed that the SAT measures aptitude rather than achievement; hence, it is much less susceptible to short-term coaching effects than to long-term, regular instruction (Black, 1979; College Entrance Examination Board, 1981). Of late, however, this situation seems to be changing. The advent of microcomputers and software packages aimed at computer-assisted instruction (CAI) to aid the student in coping with the SAT has produced some startling but mixed and inconclusive results (Biemiller, 1982; Cullen, 1982). Claims of potential increases from 50 to 175 points currently are being made following only a few hours of CAI utilizing SAT software packages (Cullen, 1982; Krell Software

Corporation, 1981). Thus, the central problem or focus of this research is that these claims need to be subjected to empirical investigation.

If the above claims be true, they suggest at least three implications. First, the SAT may be susceptible to short-term coaching effects, supporting the inference that it measures achievement rather than (or in addition to) aptitude (Coughlin, 1982). Second, college entrance standards would of necessity change, causing the admissions and scholarship officer to wonder whether the elevated score represents true "aptitude" or some form of short-lived, "artificial" intelligence. Third, not all students who seek college admission can afford the expense of CAI aimed at score elevation, creating a disadvantage for them and prompting feelings of discrimination. The implications, and possible consequences, of one more form of unequal treatment could be extensive. Hence, it seemed imperative that the reported claims be explored.

Purpose of the Study

The primary purpose of this study was to assess empirically the effectiveness of selected software in raising SAT scores when utilized with short-term CAI on the microcomputer. Additionally, the relationship of several variables (viz., the student's cognitive or learning style, socio-economic status of the family, educational level of the parents, occupation of the parents, number of siblings, gender, I.Q., high school G.P.A., and the student's proposed college

major) to SAT score gains following the controlled CAI were investigated.

Significance of the Study

With the advent of the microcomputer and the rapid expansion of its application has come a keen interest in its use in education. The storage and retrieval of student records data and its use in CAI, particularly in math labs, are but two of the more common applications. One of the principal restraints to an even broader application, however, has been the limited availability of software appropriate to the instructional process. Additionally, we have experienced a period in time when teachers felt intimidated by the computer. Some teachers have even struggled with latent fears of possible displacement. Today, however, with its pervasive presence and rapidly expanding applicability, the microcomputer has fairly wide acceptance and a viable place in the educational setting and process.

One of the more recent innovations in the use of the microcomputer has been its application in the process of tutoring students who are preparing themselves to take college admissions tests, particularly the SAT. Several software packages have recently become available which purport to afford significant gains in students' scores after only a few hours of exposure to CAI. The two most extensive and promising of these packages, primarily because of their diagnostic and strategy aspects, are available through Krell

Software Corporation and The Psychological Corporation. Unfortunately, however, to date the validity of neither of these packages has been tested empirically. To quote Cullen, "Reliable tests of personal computer software for SATs and similar exams just haven't been performed. At the moment there isn't much objective evidence to go on; SAT software hasn't been around long enough for quantitative evaluation" (Cullen, 1982, 121-122). Biemiller believed, however, that "Microcomputers will alter standardized admissions and placement tests radically in the next few years ...." (Biemiller, 1982, p. 4). Already, based upon preliminary, limited data, Friedland of Krell Software Corporation, has claimed notable gains in SAT score improvement following student use of the 81/82 Preparatory Series (Krell Software Corporation, 1981). That there is a significant potential impact upon student SAT performance seems apparent from these early impressions. The full implications, however, are yet to be observed and explored.

Conceptual Framework

Skinner's (1953) theoretical model of operant conditioning has provided the primary conceptual framework for this study. Operant conditioning, as contrasted with Pavlovian or classical conditioning, entails the development or emergence of a particular operation or pattern of behavior, known as the operant, under the influence of positive reinforcement. That is, positive reinforcement, such as food or some other prized reward, helps to shape the

desired behavior.

The typical CAI software package is built upon Skinnerian principles of operant conditioning. Specifically, there is stimulus presentation, subject response, immediate feedback relative to accuracy of the response, and positive reinforcement where appropriate. In effect, these packages, and the use of the microcomputer, are sophisticated technical extensions of the much more cumbersome Skinnerian teaching machines of the late '50s and early '60s.

In the present instance, the desired operant is to teach the student to make the correct choice on the SAT. The microcomputer creates the learning situation and provides the options. When the student makes the correct choice, positive reinforcement is provided through musical sounds and the student seeing on the monitor screen such statements as: "Good choice; you got the right answer." This process is repeated until the student learns the correct answer to each sample item.

As a secondary conceptual framework, Witkin's concepts in personality differentiation as related to cognitive style (i.e., field-dependence versus field-independence) additionally have been utilized (Witkin et al., 1974; Witkin et al., 1975). Cross (1976) has provided perhaps the clearest description of this dimension. She stated:

> The field-independent person is likely to deal

> with elements independent of their (sic, his/her) background, whereas the field-dependent person deals with the total field or situation. The field-independent person consistently approaches a wide variety of tasks and situations in an analytical way, separating elements from background. The field-dependent individual approaches situations in a global way, seeing the whole instead of the parts. (pp. 116-117)

According to Messick (1970), Kogan (1971), and Cross (1976), the field-dependent/independent dimension has been the most extensively researched among the twenty or so dimensions identified as representing cognitive style. In essence, persons differ in their perception of the world and seek to cope with it in their own unique, personal ways. Ambiguity, in particular, often facilitates this process and leads to the emergence of one's dominant style. Personality traits, individual differences, one's choice of defense mechanisms, and the dominant pattern of socially adaptive behaviors all serve to reflect a given individual's cognitive style (Witkin & Moore, 1974). In a learning setting, especially as related to information processing, one's cognitive style impacts upon outcomes (Cross, 1976). Hence, the present study incorporated this variable, as reflected in the Witkin's field-dependent versus field-independent dimension, and sought to assess its relevance in terms of each student's SAT post-test level of achievement.

Assumptions

1. The student had an opportunity to learn during the CAI or treatment portion of this study.

2. Favorable conditions conducive to learning existed.

3. The student was motivated to learn.

4. There was feedback relative to student effort.

5. There was appropriate and adequate reinforcement relative to effort.

Limitations

1. This study was confined to the use of a selected software package and to the performance of students whose representativeness is restricted. There was no systematic attempt at randomness in student identification; and all who participated did so voluntarily. Moreover, it should be stated at this point that all district requirements relative to the participation of human subjects were satisfied prior to the initiation of this study.

2. Due to the number of students involved, and to the use of a varied testing schedule, the lapsed time between pre- and post-testing did not remain constant for all students. In view of this intervening variable, consideration was given to analyzing the post-test data both with and without the use of an adjustment factor. This factor for lapsed time was taken from the literature based upon studies reflecting typical SAT score changes where non-intervention

strategies were utilized (Alderman & Powers, 1980; Messick & Jungeblut, 1981).

3. The sample size also represented a limitation, particularly where demographic variables were concerned. While the sample size was entirely adequate relative to the primary purpose of this study, the examiner was aware of the restraints imposed upon statistical analysis when a broadly classified set of variables is being considered. Nonetheless, the findings answered adequately the basic question of SAT score sensitivity to short-term CAI in a controlled, but not atypical, setting.

Hypotheses

Several hypotheses were formulated relative to the empirical questions embodied in this study. They differed in perceived importance; and most related to conditions applicable to the Experimental group. All were stated in the null or "no difference" form and tested statistically at the .05 level of confidence.

1. No difference will be evidenced in Group 1 (Experimental) and Group 2 (Control) SAT pre- and post-test scores as a consequence of CAI.

2. No difference will be evidenced in students' (classified as field-dependent versus field-independent based upon the Witkin Group Embedded Figures Test) SAT post-test scores when separated as being above or below the group mean.

3. No difference will be evidenced in students' (classified as

being of high or low socio-economic status relative to the group's median family income) SAT post-test scores as being above or below the group mean.

4. No gender-related difference will be evidenced in SAT post-test scores.

5. No I.Q.-related difference will be evidenced in SAT post-test scores.

6. No G.P.A.-related difference will be evidenced in SAT post-test scores.

7. No difference will be evidenced in SAT post-test scores when separated relative to educational attainment of the parents.

8. No difference will be evidenced in SAT post-test scores when separated relative to occupational status of the parents.

9. No difference will be evidenced in SAT post-test scores when separated relative to students' number of siblings.

10. No difference will be evidenced in SAT post-test scores when separated relative to each student's proposed college major.

CHAPTER II

REVIEW OF RELATED LITERATURE

The relevant literature in this area of study was separated into seven segments: 1) the historical use of college admission scores, 2) the effects of tutoring upon SAT scores, 3) attacks upon the SAT, 4) defense of the SAT, 5) potential impact of the microcomputer, 6) computer-assisted instruction, and 7) cognitive styles.

Historical Use of College Admission Scores

Requiring admissions testing is a common practice of colleges and universities throughout this country. The purpose of such testing is to determine whether the prospective student meets specific requirements as to ability and achievement potential. While performance on these tests is important in setting and maintaining standards, test scores per se are not the sole criterion for admission. Outstanding high school grades, a report of participation in events apart from academics (i.e., social and/or athletic), evidence of community service, and personal recommendations also frequently are taken into consideration. Thus, it may be stated that minimal test scores are necessary for admission, but they are not sufficient in themselves (Astin, 1965).

There is not now, nor has there ever been, one universal test used for college admission in this country. From 1901 until the late '20s, a series of tests developed by the College Entrance Examination Board was utilized rather extensively (Holloway, 1971). These instruments were principally area tests that dealt with the core curriculum (i.e., history, physics, chemistry, mathematics, and languages). Holloway (1971) wrote:

> In the late 1920s Carl Brigham offered a new tool (the SAT) for measuring the college potential of students; and aptitude testing began to exercise a liberating effect on the process of student selection, enabling colleges to supplement school records and to evaluate with more accuracy candidates from all sections of the country. (p. 213)

Until 1959 the SAT was, in effect, as close to a sole instrument for college admissions screening as ever has existed in this country (Elliott, 1978). At that time, however, the American College Testing Program was initiated; and the ACT (American College Test) was introduced as a rival instrument. To date, the SAT and the ACT have remained as the principal college admissions screening instruments utilized in this country.

Robert J. Kingston (College Entrance Examination Board, 1979), past-president of the College Board, reported:

> The SAT is designed to be a test of students'

> reasoning abilities in verbal and mathematical areas, in order to provide colleges with a common denominator by which to estimate the likely college performance of students from all types of schools, from all parts of the country. It is not a measure of anyone's individual worth or of a school's effectiveness. And it is not the most important factor in admissions. On the other hand, it is as good and valid a predictor of academic success during a student's first year of college as it has ever been. (p. 6)

At present the SAT remains the primary, most widely used college admissions screening instrument available in this country, though the ACT is growing in strength each year. While many colleges will accept scores from either instrument, some accept only one or the other (most generally the SAT); and a few require both.

Effects of Tutoring upon SAT Scores

Alderman and Powers (1980) reported on the effectiveness of special preparation programs for improving SAT-Verbal scores at the high school level. The traditional lecture-tutorial approach resulted in a gain of experimental subjects over controls amounting, on the average, to about one additional correct item. Black (1979) asserted that no amount of coaching will appreciably affect scores

on the SAT. Coffman (1980) concluded that special instruction in the development of problem solving skills can improve scores, although long-term, regular school instruction is thought to be more effective. Coughlin (1982) tended to agree and stressed the importance of a strong secondary curriculum and the need for serious student attitudes toward scholarship as being the clearest factors affecting SAT scores. Farr and Olshavsky (1980) concurred, stating that the recent emphasis upon minimal competency in the basic skills at the secondary level will have little effect in raising SAT scores.

Reports by the Federal Trade Commission (1978; 1981) concluded that some commercial coaching schools have produced statistically significant score gains (viz., typically around twenty-five points, or between two and three extra items correct) but that these gains are modest and of limited practical consequence. Traditional coaching efforts are just not worth the returns, according to Gallagher (1980). Moreover, Hillje (1980) found that the development of basic reading and writing skills, rather than the mastery of long vocabulary lists, has proven helpful in minimal score improvement. Messick and Jungeblut (1981) reported that time and method of instruction (i.e., long-term, regular instruction of high quality and standards), individual ability, and student motivation are the principal factors affecting score changes. Powers (1981) reported that students perceive coaching efforts as being effective and worth the time and expense involved in spite of the minimal gains. Thompson and DeLeonibus (1978) found, in the face of national test score

declines, that a strong basic curriculum in English and math, plus an attitude of academic rigor, appear to be of principal significance in maintaining SAT scores (at selected secondary schools). Finally, but perhaps of lesser importance, Gifford and Fluitt (1980), Kubistant (1981), and the editors of Career World ("Will Coaching," 1984) reported that the development of test-taking skills and repeating the test in practice sessions have the positive effects of reducing anxiety and building confidence.

Attacks upon the SAT

Due in part to frustrations experienced in attempts to manipulate SAT scores, a number of authors have attacked the test as a poor predictor of academic success in college. Boyd (1980), in particular, said that the SAT is especially poor where black students with good scholastic backgrounds are concerned in that these students tend to out-perform predicted expectations. Agreeing with Boyd, Dalton (1974) concluded that high school achievement (G.P.A.) is a better predictor of college potential than is the SAT where minority students are concerned. Bracey (1980) questioned the predictive validity of the test and called for newer and more relevant screening methods commensurate with changing societal needs. In two separate articles, Jencks and Crouse (1982) argued that the SAT measures achievement rather than aptitude, that standardized achievement tests are better screening instruments for college admission, that SAT items do not correlate well with typical high school curricula,

and that SAT scores do not reflect or reward high achievement at the secondary level. Slack and Porter (1980) and Ryor (1977) agreed with Jencks and Crouse and recommended that the SAT be dropped as a college admissions requirement.

Defense of the SAT

As one might anticipate, there is no shortage of authors who come quickly to the defense of the SAT. Angoff (1973-74) noted that the SAT's predictive validity (for white, middle-class students) and its reliability are strong points. Burns (1976) stressed an increase in minority participation and a growing discrepancy between test objectives and secondary instructional objectives as primary causes for SAT score declines. Edmonds (1982) responded to criticisms by Jencks and Crouse, stating that their attack upon the SAT is unwarranted and that it only contributes to confusion and negative attitudes in society relative to test accuracy and effectiveness in screening for college admission. Jackson (1980) responded to Slack and Porter, attempting to defend the predictive validity of the test. Karney (1980) laid the blame for declining scores upon poor discipline in the public schools. Messick (1981) reviewed the research findings to date relative to SAT coaching and reinforced Educatonal Testing Service's position that aptitude tests such as the SAT are not susceptible to the effects of short-term coaching efforts.

Potential Impact of the Microcomputer

The potential impact of the microcomputer and its related software has already been discussed earlier in this paper. As previously referenced in Significance of the Study (on page 3), the opinion of several authors and selected publishers is that the microcomputer holds the promise of quite positive SAT score gains following even short-term CAI. Friedland (Krell Software Corporation, 1981), in particular, reported anticipating as much as 175 point increases when students use the company's preparatory series. No other software publisher to date has made these kinds of claims. Moreover, such claims are not consistent with the results of earlier coaching efforts utilizing more traditional methods.

Prior to the advent of the microcomputer and its related software, tutorial assistance aimed at raising students' SAT scores focused upon lectures, reading assignments, a combination of the two, and practice sessions for improving test-taking skills. The latter method was intended to reduce anxiety and to strengthen student confidence. To a lesser extent emphasis additionally was placed upon: 1) providing diagnostic assistance relative to individual strengths and weaknesses, and 2) offering limited strategies for coping with the test. None of these methods was very effective, however, resulting in only modest score gains. While on occasion statistical significance was achieved, the practical consequences of the reported gains were minimal. Hence, the question yet remains open relative to the overall effectiveness of intervention strategies

aimed at improving SAT scores.

Some are of the opinion, however, that the microcomputer software and the related CAI might not be as promising as earlier anticipated. As recent as September, 1984 Stanley H. Kaplan, prominent educator and founder of the Stanley H. Kaplan Educational Centers, was interviewed in Career World where he was asked the following questions: "What is your opinion of the new software test-preparation programs for the SAT? Do you use computer-assisted teaching techniques?" His reply: "So far it's not practical. It's electronic page-turning. There is no real input by the student. Cramming, whether by computer or otherwise, means trying to squeeze an impossible amount of knowledge into an impossibly short period of time" ("An interview," 1984, p. 18). The literature has stated for years that long-term basic instruction in a rigorous curriculum is the only effective way to improve SAT scores. The new technology is attempting to change this; and the present study is in part prompted by such interest.

Computer-Assisted Instruction

It should first be stated that computer-assisted instruction was by no means the focal point of this study. Rather, CAI was simply the instructional medium or vehicle through which the SAT software was utilized and assessed. Due to the relevance of computer-assisted instruction, however, it seemed appropriate that a brief historical overview be included in this review.

Computer-assisted instruction made its initial appearance in the early 1950s, utilized principally in the training of new personnel in the computer industry. Shortly thereafter, Stanford University initiated a tutorial program aimed at teaching elementary school children skills in mathematical logic (Suppes & Macken, 1978). The program grew in scope, began incorporating the teaching of reading skills, and later became known as the Brentwood Project (Atkinson & Wilson, 1969). Other universities took Stanford's lead; so by 1970 Illinois, Texas, Brigham Young, Pittsburgh, and Florida State had active tutorial programs (Blaisdell, 1976-77).

According to Nievergelt (1975), "Computer assisted instruction during the decade of the '60s was characterized by a number of limiting factors: insufficiently powerful computers and terminals, restriction to a few rigid teaching strategies, and the splitting of resources among too many projects below critical size" (p. 465). During the 1970s, however, CAI underwent remarkable changes, that is as seen in such large-scale projects as PLATO and TICCIT (Bunderson, 1972; Nievergelt, 1975; Magidson, 1977). Of similar importance, additionally, were such projects as LOGO, started by Feurzeig and Papert (Nievergelt, 1975), and SOLO at the University of Pittsburgh (Dwyer, 1972).

The work of Skinner at Harvard, especially his programmed instruction movement utilizing teaching machines, would most generally be considered as having been highly influential in shaping CAI developments (Skinner, 1954; Blaisdell, 1976-77). Of special significance

was his utilization of branching techniques in the instructional mode. These techniques permit movement into different areas through the introduction of new content or the utilization of additional materials in an area of apparent learner weakness, while continuing to move toward a previously determined instructional goal. Utilization of this technique is also a basic element of the instructional methodlogy employed in much of the software programming common to CAI today.

Also related to the present inquiry is a report by Lockhart (1981) wherein she reviewed the military research on the effectiveness of CAI under replicated and retention conditions. Of special interest was a series of studies conducted at the Great Lakes Naval Training Center in which test item response format was found to relate significantly to retention rates. That is, CAI which was later assessed under a recall, rather than recognition, response format was found to be more stable and enduring. These findings are recognized as falling well within the classical body of psychological research involving learning and test item response format. Herein it has long been known that learning which transpires under recall conditions has a higher or longer retention rate than that which occurs under recognition conditions. In this respect, at least in the reported instance, learning which took place under a CAI methodology generated results entirely consistent with classical findings.

In the present era there appears to be little doubt that CAI has made notable gains in the educational community. Funds for the acquisition of both hardware and software have more of late become

increasingly available; and there is a growing emphasis upon the need for a broader spectrum of computer literacy. Curriculum changes at the elementary, secondary, and higher education levels are reflecting the recognition of this need and the public's demand that it be met. The present study is clearly in the mainstream of this trend.

Cognitive Styles

Allport (1961) probably was the first to identify individuals as having unique "styles" of learning (similar to certain personality "traits"); and Witkin and his associates have researched the concept extensively. Most recently, Cross has been one of the more prolific writers in the field. In Accent on Learning (1976), she stated:

> It takes no special knowledge of research to recognize that we all have characteristic "styles" for collecting and organizing information into useful knowledge. Some people do their best learning in interaction with others, some in lone study and contemplation. Some people prefer to learn a skill by manipulating concrete objects, some by watching, some by listening, some by reading the manual of instructions. Some people approach learning tasks systematically and methodically; others are more intuitive and global. In brief, people

> have characteristic ways of using their minds. Psychologists label these mental characteristics cognitive styles, and their importance to education is fundamental and pervasive. (p. 112)

Cross stated further that at least a dozen or so cognitive dimensions have been extensively researched to date and that an equal number or more have been identified but not yet researched. Among the more common ones, Witkin's bipolar dimension of field-dependence versus field-independence has thus far claimed the greatest attention. It is this specific dimension which was focused upon in the present study and which is discussed in greater detail at a later point in this paper.

Summary

A review of the related literature has encompassed seven topics. Four of these dealt specifically with the SAT, ranging from its use as a college admissions screening instrument to the effectiveness of attempts to alter scores through a morass of coaching methodologies. Growing out of these attempts has come a variety of attacks upon, and defenses of, the instrument relative to one's posture as regards its usefulness for its intended purpose. There is no limit to proponents on either side. In the present instance, the chosen methodology focused upon the use of the microcomputer in assessing the effectiveness of selected SAT software intended to

raise scores. While this approach is uniquely contemporary, and is claimed by some to be of promising potential, the verdict is not yet in since adequate data have thus far not been reported in the literature. The present study aims to help fill this void. Consideration has herein also been given to the relevance of this methodology as it impacts upon individual students' learning styles, with Witkin's field-dependence versus field-independence being the principal dimension under review. In essence, this overview of the literature has concerned itself with the instrument, the teaching methodologies utilized, and the impact of the latter upon individual student learning.

CHAPTER III

METHODOLOGY

Population

Forty-eight college-bound high school juniors in Escambia County (Florida) served as subjects in this study. They were invited to participate on a voluntary basis from a pre-identified group provided by guidance counselors throughout the district. All students who volunteered participated in SAT pre-testing. Based upon the rank order position of these scores, each student was assigned to either the Experimental or Control group via an AB, BA sequence. Twenty-four subjects were assigned to each group. They were advised that full participation was a part of the contractual obligation and that all, in time, would be offered computer-assisted instruction.

Initially sixty-five students were identified and contacted to appear for testing. However, on the actual day of SAT pre-testing only forty-eight arrived and took all three tests (viz., the SAT, the Otis-Lennon School Ability Test, and the Witkin Group Embedded Figures Test). It was later learned that twelve students from one school and three from another were away on band trips; and two more were on family vacations. Hence, these students were not able to participate and were excluded from the

potential sample. Among the forty-eight who did participate, however, five withdrew prior to completion of the fifteen hours of CAI. Two were in the Experimental group, and three were in the Control group. In each instance these students withdrew in order to accept part-time work assignments which eventually led to full-time summer jobs. Their primary motivation was to accept employment opportunities prior to the advent of much heavier competition at the close of school. These new responsibilities precluded their having sufficient extra time to continue participating in this study. Hence, the total post-CAI sample for data analysis was reduced to forty-three.

Nonetheless, the sample size was entirely adequate for the primary purpose of this study. Where the demographic variables under consideration are concerned, however, two things should be noted: 1) the effects of these variables are normally studied with large samples; and 2) the results reported in the literature refer to a cross section of average students and not to highly selected students as have herein participated. In particular, the mean and/or median socio-economic status of the family is most often considerably lower, as are the educational level and occupational status of the parents. In essence, these students are not typical or "average" SAT takers. They were accepted, however, because they are bright, highly successful achievers who were well-motivated to complete a voluntary study aimed at testing the effectiveness of a learning methodology intended for their benefit.

It should additionally be stated at this time that the matter of sample size was taken into consideration at all points deemed appropriate in the data analysis. The rationale for this decision rested prinicpally upon two factors: 1) a response to the basic question of what the analysis would reveal if the sample size were doubled, and 2) the fact that the replicated portion of this study in Phase II permitted a logical combination of Groups 1 and 2 following receipt of the CAI. As will be noted in the various reports of these combined analyses, at no point did the outcome alter earlier inferences based upon a comparison of the performances of Groups 1 and 2 treated separately.

Data Collection

Design

This study utilized a classical pre- and post-test two-phased design, with the SAT being the primary testing instrument. Phase I consisted of: 1) pre-testing of all subjects, 2) fifteen hours of CAI for the Experimentals, and 3) post-testing of all subjects three weeks later. Phase II consisted of: 1) fifteen hours of CAI for the Controls, and 2) a second post-testing of all subjects. Utilization of this design provided: 1) a pre- and post-test analysis of Experimental subjects' score changes as gains or losses following CAI, 2) a pre- and post-test analysis of Control subjects' score changes as gains or losses without CAI, 3) a baseline of comparison when Experimental subjects' post-test score changes with CAI were

evaluated relative to Control subjects' post-test score changes without CAI, 4) a retention measure for Experimental subjects' score changes following CAI when the results of the two post-tests were compared, and 5) a replication analysis of Control subjects' score changes as gains or losses following CAI.

Additionally, certain conditions prevailed. First, subjects were expected to complete all coaching sessions. Second, attempts at preparation outside of the controlled sessions were specifically prohibited. Third, the retention period for the Experimentals was set as the lapsed time between the two post-tests. This was in part controlled by the fact that Educational Testing Service's SAT testing sessions are offered only in the spring and fall of each year. Consequently, time constraints necessitated utilizing the final session of the spring. Moreover, while the results of briefer or longer retention periods might certainly be of interest, such is typically determined by the researcher and clearly set forth within the design. This did occur in the present study.

Further, two points appear to be of particular interest. First, all students completed their last few hours of CAI during the two days prior to the SAT post-test, with over ninety percent of them completing it on the last day. Second, all students were tested at the same time during the three separate testing dates. There was, however, a difference in lapsed time between pre- and post-testing following CAI for Groups 1 and 2. Specifically, twenty-one days elapsed for Group 1; and twenty-eight days elapsed for Group 2,

resulting in a difference of only seven days. This researcher has found nothing in the literature to suggest that a seven day lapsed time between pre- and post-test comparisons involving the SAT performances of two equal but separate groups would have any notable effect upon the results. Take, in particular, the reports of Alderman and Powers (1980), Messick and Jungeblut (1981), and the Federal Trade Commission (1978; 1981), wherein average gains of only one to three extra items correct occurred following weeks to months of coaching. Even beyond this, the general instability of SAT scores from one testing session to the next would tend to preclude the necessity of a separate analysis in this instance and clearly would seem to represent a matter of greater concern (which has been dealt with under Findings of the Study). Hence, there appears to be nothing to justify the use of an adjustment factor (referred to under limitation two on page 7) in the analysis of these comparative data.

Equipment and Materials

The equipment for this study consisted of three sets of Krell SAT software and ten Apple IIe microcomputers. Krell Software Corporation provided the software packages at no cost to the examiner. Containing six disks each, these materials were in sufficient quantity to allow each subject to work independently. The disks in these software packages contained eight separate menus, with each consisting of 1,000 questions or problems. There were four disks

for the Verbal area and two disks for Math. The vocabulary disk alone contained over 32,000 entries, with thousands of additional words in the reading passages and written problems of Math. No student was able to exhaust all of these entries during the fifteen hours of CAI, permitting extensive and almost unlimited review items for areas of weakness. Emphasizing the need to work in these specific areas was the primary orientation and focus at all times during this study. In addition, the examiner provided three dictionaries for word study, plus paper and pencils for computations in Math.

The microcomputers were available in the administrators' computer lab at West Pensacola Elementary School, which is centrally located within the district. The lab was available for CAI each day after school (from 3:00 p.m.) until as late as students wished to stay and all day on the weekends during the period of this research. This schedule permitted a total of forty-eight days for completion of all CAI and the three separate testing sessions. Students came to the lab at their own convenience, signed in and out on an individualized, continuous log, and studied as long as they desired. Most stayed between two and three hours on each visit. The lab was supervised at all times by at least one of four adults; and a quiet, serious atmosphere was maintained. Students worked alone at individual monitors, and tutorial assistance was not offered. The learning mode was strictly CAI.

Comforts and conveniences at the lab were outstanding. The room was large and well lighted, with carpet on the floor. There were no windows; the building was centrally air-conditioned. Restrooms were conveniently situated across the hall; and a sitting room with a private telephone for calling in or out was also available. Female students were especially appreciative of this for maintaining contact with home during evening and weekend hours. In addition, a large one-way viewing mirror in an adjacent room permitted observation of the entire lab by visiting parents. Lastly, two large refrigerators were available for holding lunches and cold drinks while students studied. The situation could not have been more conducive to creating a relaxed but serious and secure learning environment. Consequently, the students were most cooperative and responded in a very appreciative manner.

Instruments

Several different instruments were utilized in this study. The SAT has been mentioned repeatedly as the primary instrument. It was administered by the Testing Center staff on regularly scheduled dates at Pensacola Junior College. The dates utilized were set by Educational Testing Service (ETS) as part of its officially sanctioned, nation-wide testing program. Pre- and post-test sessions for this study were determined per the dates offered.

The Otis-Lennon School Ability Test (O-LSAT) was administered to all subjects to determine each student's level of intellectual

functioning (I.Q.). The acquisition of these data was related to the common inference that I.Q. is significantly correlated with achievement. Moreover, it was particularly important to know that the two groups involved in this study were equally matched relative to general ability.

The Witkin GEFT was also utilized (Oltman et al., 1971; Witkin et al., 1971). Each student was identified as field-dependent or field-independent. Predictions were made, based upon these classifications, relative to each student's anticipated performance (i.e., as a high achiever or low achiever) on the SAT post-tests.

A brief questionnaire was utilized to obtain demographic data for each student. One specific bit of information sought involved family income. These data were relevant in that socio-economic status has been shown to relate significantly to SAT performance (Burns, 1976; Dalton, 1974).

Information was also obtained relative to each student's G.P.A. in the district's core curriculum. These data, along with the SAT pre-test scores, provided a rough estimate of both ability and motivation to achieve.

Gender has been shown to be related to differential levels of performance on the SAT (Messick & Jungeblut, 1981). Hence, such related differences in subsequent performance on the SAT post-tests were examined following application of the CAI to assess their significance under these conditions.

Data relative to the educational level of the subject's parents were obtained and classified into three groups: a) graduate or professional school, b) college, and c) high school or below. Post-test SAT scores were separated to reflect "high" and "low" achievers. Both sets of data were then placed in a contingency table for statistical analysis.

Occupation of the parents was treated similarly. Classifications were made according to the following: a) professional, b) white collar, non-professional, and c) blue collar. Comparisons with SAT post-test scores via a contingency table were then made.

Family income was classified as follows: a) above the group median, and b) below the group median. Comparisons with SAT post-test scores via a contingency table were again made.

Number of siblings lent itself readily to tabular presentation and analysis. The data were placed in a contingency table as follows: a) zero siblings, b) one sibling, and c) two or more. The data determined these classifications and the related SAT analysis.

The student's proposed college major was classified and treated as follows: a) math, science, or engineering; b) fine arts, social studies, or humanities; and c) vocational-technical or business related. Analyses similar to the foregoing were applied.

Data Analysis

Nonparametric techniques (Siegel, 1956) were utilized in all statistical analyses for three primary reasons: 1) the sample permitting group comparisons was relatively small, though adequate; 2) the data represented ordinal scaling at best; and 3) irregular distributions were anticipated. Moreover, in each instance the commonly used probability value of .05 was set as the minimally acceptable level of confidence for statistical significance.

The Mann-Whitney U Test was utilized extensively in the analysis of data obtained in this study. More specifically, it was applied in the following instances: 1) to determine on the basis of SAT pre-test scores that the Experimental and Control groups did not differ significantly in terms of student aptitude and motivation to achieve; 2) to draw similar inferences relative to the results of the O-LSAT; 3) to assess the significance of gender-related SAT pre- and post-test performance differences; 4) to assess the significance of I.Q.-related (O-LSAT) SAT post-test performance differences; 5) to assess the significance of G.P.A.-related SAT post-test performance differences; 6) to assess the significance of SAT post-test gains (i.e., relative achievement) of members of the Experimental group under CAI treatment conditions versus those attained by Control group members without treatment; and 7) to assess the significance of SAT post-test gains as a function of the CAI treatment applied to students when they are separated into high and low groups relative to socio-economic status. The inference here was that high-status

achievers, as a consequence of the CAI treatment, would significantly out-perform low-status achievers.

It should be noted at this point that item 6 above was considered the heart of this study. All other aspects were thought to be supplemental and supportive. In this regard, detailed analyses of both group and individual score changes have been provided in a variety of formats. Tabular and graphic presentations have afforded the most succinct and clearest pictures of these score changes. Perhaps most revealing, especially as related to the basic empirical questions regarding Krell software and short-term CAI effectiveness, are the portions of this analysis which deal with the concept of the standard error of measurement. SAT score instability under repeated testing conditions was treated relative to this concept, with that analysis being thought to have provided new insights into this whole area of investigation. Hence, it should be stated that reported SAT score changes, under any coaching methodology, must be evaluated relative to this concept. When such occurs, one might find that even the most highly acclaimed score gains fade in significance.

Experimental subjects were classified as high and low achievers relative to the group mean. Based upon their SAT post-test scores, a two-by-two contingency table was constructed; and these scores were then placed in the appropriate cells relative to each student's level of achievement and previous classification as being field-dependent versus field-independent. A chi-square test was herein applied to these data to determine their statistical significance.

Gender-related SAT post-test performance differences were treated similarly in a two-by-two contingency table. One axis of the table reflected gender, while the other indicated high or low achievement. A chi-square test was applied to assess the significance of observed differences in relative performance as a function of the CAI treatment. The point of interest here was to appraise the possible impact of CAI on eliminating traditionally observed differential achievement. It should be noted, however, that this particular analysis was based upon the anticipation that SAT pre-test scores evidenced a gender-related performance difference. Otherwise, the analysis would have relevance only should it demonstrate a significant change in performance scores in the opposite or unanticipated direction.

Chi-square contingency tables were also used in the analysis of data related to the educational level and occupational status of the parents, family income, number of siblings, and each student's proposed college major. The intent here was to assess the impact or relevance of these demographic variables as they related to SAT post-test score gains.

A correlational matrix involving Spearman Rho coefficients was also generated. These data reflect correlations between SAT post-test scores and the following: 1) O-LSAT scores, 2) G.P.A., and 3) socio-economic status. The primary intent was to assess the predictive value of these measures relative to SAT post-test performance. Similarities among these predictors were also evaluated.

Data from the above analyses have been summarized and reported in tabular and graphic form. Included as well are notations of relationships or differences reflecting statistical significance. These findings, and their possible implications relative to the empirical hypotheses previously set forth, are discussed in narrative form. Conclusions and inferences are offered as deemed appropriate.

CHAPTER IV

FINDINGS OF THE STUDY

Representativeness of the Sample

Prior to reporting the findings of this study, it would seem appropriate first to answer the following question: How well does the sample, which generated these findings, represent the larger population from which it was drawn? When voluntary subjects are used, this is a particularly important question. It should be stated initially that five out of six of the large high schools in the Escambia School District are represented in the sample. Students from one additional large school, and two small rural schools in the extreme north end of the county, were not invited to participate due to the traveling distance required for receipt of the CAI at the computer lab. There is no reason to believe, however, that these particular students would have performed any differently than did those who participated. In fact, test scores from all sources, including the SAT, presently on file in the Guidance Department (E. Parker, personal communication, August, 1984) and Evaluation Services have not shown themselves to be notably different in any respect. Consequently, it seems reasonable to assume that the sample obtained represents the predicted performance of comparable students at each of these schools.

Impressions drawn from data provided by previous students need not stand alone, however. The data from this study were evaluated by means of a chi-square goodness-of-fit test (Table 1); and the results of this analysis reinforce the above inferences. Specifically, the obtained chi-square value was 5.17; while the critical value was 11.07 (Guilford, 1956, p. 540). Based upon the analysis of SAT Total pre-test scores, it may be concluded that the sample is normally distributed, representative of the larger population from which it was drawn, and adequate for the purposes of this study.

Given that the SAT Total pre-test scores for the entire sample are adequate as evaluated and presented in Table 1, the next logical step was the systematic assignment of students to Groups 1 and 2. Students in Group 1 served as Experimentals; while those in Group 2 served as Controls. Group assignment was based upon the following procedure: 1) rank all SAT pre-test scores from highest to lowest; and 2) assign each student to his/her respective group by means of an AB, BA sequence, beginning with the initial assignment to Group 1. In the not uncommon situation where tied scores occurred, students were assigned in alphabetical order, from which point the process was continued per the above procedure. Under this arrangement all students were assigned such that matched pairs and equated groups resulted. This plan lent itself readily to a study of the impact of treatment (CAI) applied to one group (Experimentals) as compared to a second group (Controls) which did not receive treatment. This comparison formed the principal basis of the study.

Table 1

Chi-square Goodness-of-Fit Test Applied to SAT Total Pre-test Scores to Demonstrate that the Scores Are Normally Distributed and Representative of the Larger Population

Score range	n	Mean	Std. dev.	Crit. value	χ^2	|p|
770-1290	48	1025	126	11.07	5.17	>.05

Legend:

Score range = SAT Total pre-test scores, ranging from a low of 770 to a high of 1290.

n = the total number of students in the sample.

Mean = the "average" SAT score in the pre-test sample.

Std. dev. = the standard deviation, which is a measure of dispersion indicating the extent to which students' scores vary about the mean.

Crit. value = Critical value, the minimum chi-square value required for statistical significance.

χ^2 = the chi-square value calculated in the goodness-of-fit test.

|p| = the probability value or statement related to statistical significance.

In order to show that matched pairs and equated groups did in fact result, the Mann-Whitney U Test was applied to SAT Total pre-test scores. The outcome of this analysis is reflected in Table 2. As is evident, the total ranking (R) of scores for each group differs very little (i.e., 589.5 for Experimentals and 586.5 for Controls); and the probability of occurrence of such a small difference would be .4880, or approximately 49 times in 100 through repeated sampling. Thus, the null hypothesis of "no difference" between the SAT Total pre-test scores of the two groups of this sample must be accepted, thereby supporting the identification of matched pairs of students and the inference of two stochastically equal groups. Stated more directly, the Experimental and Control groups did not differ prior to treatment (CAI) intervention.

The next point of interest related to the possible impact of the variable of intelligence (I.Q.) upon the pre-test scores. In other words, did students in the Experimental and Control groups differ significantly in I.Q., as reflected in the scores on the O-LSAT? Application of the Mann-Whitney U Test (Table 3) to these scores revealed two things: 1) students in the Control group are slightly brighter than the Experimentals as a group, and 2) this difference is not statistically significant and would be expected to occur through repeated sampling approximately 37 times out of 100. Hence, it may be concluded that general intelligence would not be a relevant variable in determining any observed group differences following CAI.

Table 2

Mann-Whitney U Applied to SAT Total Pre-test Scores When Separated by Group

Group	n	R	U	z	|p|
1-Experimental	24	589.5	286.5	-.03	.4880
2-Control	24	586.5			

Legend:

R = the sum of the ranks assigned to the SAT scores obtained by the students in each of the two respective groups. In this instance all scores are ranked from lowest to highest across groups, that is irrespective of the student's group membership.

U = the Mann-Whitney U calculated value.

z = the standard score associated with the U value.

|p| = the probability of occurrence of the calculated U value.

Table 3

Mann-Whitney U Applied to Otis-Lennon School Ability Test I.Q. Scores When Separated by Group

Group	n	I.Q. range	Mean	R	U	z	|p|
1-Experi.	24	105-142	122	572.5	272.5	-.32	.3745
2-Control	24	102-144	123	603.5			

Application of the Design

In brief summary, it has been established at this point that the sample was normally distributed and representative of the larger population of Escambia County college-bound juniors, that Groups 1 and 2 did not differ significantly on the basis of SAT Total pre-test scores, and that general intelligence would not account for observed group differences in post-test scores. Thus, it seemed appropriate to proceed with the application of the design.

The study was divided into two phases. Phase I consisted of pre-testing (which has been reported), application of the treatment variable (CAI) to Group 1, and post-testing of both Groups 1 and 2. Phase II consisted of the application of treatment to Group 2, following the first post-test, and then administration of the second post-test to both Groups 1 and 2. Utilization of this design permitted, during Phase I, a comparison of Experimentals and Controls based upon pre-test results, plus determination of the significance of any observed differences in post-test scores following CAI for the Experimentals. Additionally, Phase II permitted, through application of CAI to the Controls, a replication of the treatment portion of Phase I, plus the establishment of a retention curve for the Experimentals approximately thirty days following their CAI.

Hypothesis Testing

Hypothesis One

Hypothesis One states that no difference will be evidenced in Group 1 (Experimental) and Group 2 (Control) SAT pre- and post-test scores as a consequence of CAI. The outcome measures generated by the design outlined above are reported in Tables 4 and 5. The results noted in Table 4, from an application of the Mann-Whitney U Test comparing SAT pre- and post-test scores, indicate that in every instance (i.e., Verbal, Math, and Total) the sum of the ranks of the post-test scores (Post-R) exceeds the sum of the ranks of the pre-test scores (Pre-R). This means that the post-test scores, as a group, were higher at every point of comparison, although at no point was the difference between these sets of scores sufficiently large to be statistically significant. Thus, it must be inferred that the CAI for Group 1 (under Phase I) and Group 2 (under Phase II) was not effective in generating increased SAT scores beyond chance expectations in any of the areas tested. Moreover, combining these groups, and thereby doubling the size of the sample, yielded essentially the same results. Based upon the publisher's literature, it was anticipated that Experimental subjects' scores would improve significantly following CAI. Without such intervention, however, it was not expected that scores for the Controls would change notably; and they did not. This means that Hypothesis One, as stated, was sustained and accepted.

Table 4

Mann-Whitney U Applied to SAT Pre- and Post-test Score Changes Following CAI

Group	SAT	Pre-R	Post-R	U	z	|p|
1-Experi.	Verbal	463.5	526.5	210.5	-.74	.2296
	Math	492.5	497.5	239.5	-.06	.4761
n=22	Total	479.5	510.5	226.5	-.36	.3594
2-Control	Verbal	433.0	470.0	202.0	-.47	.3192
	Math	415.5	487.5	184.5	-.91	.1841
n=21	Total	419.0	484.0	188.0	-.82	.2061
1 & 2	Verbal	1769.5	1971.5	823.5	-.87	.1922
Combined	Math	1800.0	1941.0	854.0	-.61	.2709
n=43	Total	1778.5	1962.5	832.5	-.79	.2148

Legend:

Pre-R = the sum of the ranks assigned to the pre-test scores.

Post-R = the sum of the ranks assigned to the post-test scores.

The general picture becomes even clearer when one examines the data offered in Table 5. These data reflect outcome measures, expressed as gain and loss scores, comparison scores, and retention scores. In no instance did SAT scores, as a group, drop under the influence of CAI. Without CAI intervention, however, the Math scores for the Control group (in Phase I) dropped an average of 10 points between pre- and post-testing. In all other instances the scores, as a group, went up following CAI. This was true even during the retention period for Group 1 in Phase II. During the replicated portion of this study (i.e., Controls in Phase II), group score gains following CAI either equal or exceed those reported for the Experimentals in Phase I. On the whole, however, the observed group score changes reflect either no gain or modest gains following CAI. As indicated in Table 4, these gains are not statistically significant. What they do reflect is a very limited impact on the part of CAI when applied to Group 1, Group 2, and the two combined.

Portraying the results in Table 5 graphically, Figures 1, 2 and 3 reveal trends toward SAT group mean score improvement in both Verbal and Math following CAI intervention. These modest group gains are not beyond chance expectations, however, as earlier reported in Table 4. Nonetheless, a few impressions merit noting. In Figure 1, for instance, it is apparent that Verbal scores for the Controls did not change during Phase I, which was the period when these subjects received no CAI. Under the influence of CAI, however, Verbal scores for both the Experimentals and Controls, as a group, rose a modest

Table 5

SAT Pre- and Post-test Group Mean Score Changes Reported as Gain or Loss

		Phase I				Phase II		
		Pre-test	First Post-test			Second Post-test		
			CAI Outcome Score			Retention Score		
Group 1		Mean	Mean	Gain	Loss	Mean	Gain	Loss
Experimentals	Verbal	490	500	10	--	510	10	--
n=22	Math	530	530	0	--	560	30	--
	Total	1020	1030	10	--	1070	40	--
			Comparison Score			Replicated CAI Outcome Score		
Group 2		Mean	Mean	Gain	Loss	Mean	Gain	Loss
Controls	Verbal	500	500	0	--	510	10	--
n=21	Math	540	530	--	10	550	20	--
	Total	1040	1030	--	10	1060	30	--

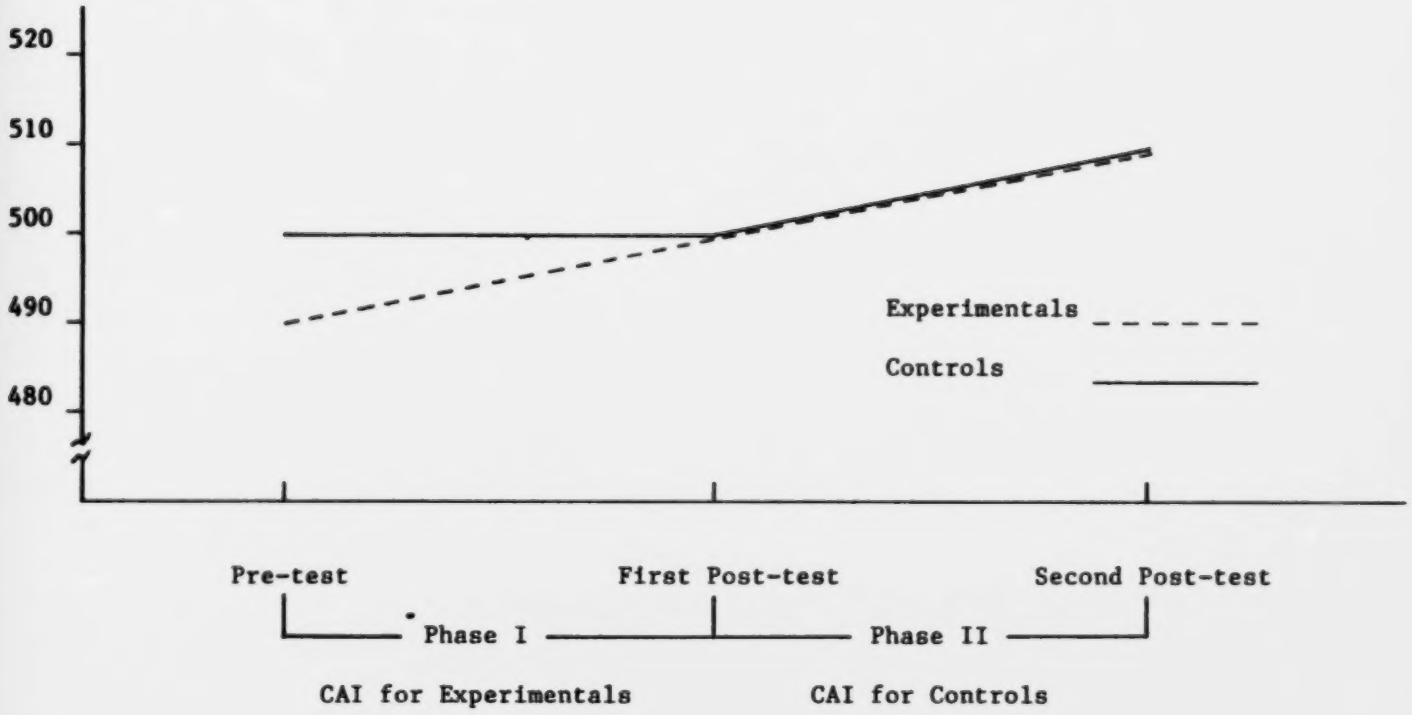

Figure 1. SAT Verbal Pre- and Post-test Group Mean Score Changes

amount. In other words, an upward trend becomes readily apparent upon the introduction of CAI.

In Figure 2 the picture for Math scores is slightly different. During Phase I (i.e., CAI for the Experimentals but nothing for the Controls), Math scores either made no progress (that is, in spite of CAI) or came down (without CAI). In Phase II (i.e., CAI for the Controls but nothing for the Experimentals), however, the score trend turned upward following CAI. This would suggest that, for whatever reasons, the subjects in both Groups 1 and 2 were better able to handle the Math sections of the second post-test than they were those of the first post-test. While CAI may have helped, it was not the only factor operating; and it clearly was not sufficient to produce results which reflected statistical significance.

Figure 3 is a compilation of Figures 1 and 2. The SAT Total results are a combination of the Verbal and Math scores, which already have been treated separately. As one would anticipate, the general trends are clearly reflective of those already reported as regards the Verbal and Math subtests. The only difference noted is that the increase in the growth rate of the curves is a little sharper due to the combining of the two sets of scores to form the Total score. Beyond this, however, it might also be noted that the CAI apparently was slightly more effective for Group 2 subjects than it was for subjects in Group 1. By contrast, the Group 1 subjects continued to improve their post-test scores (following CAI) without the continuation of CAI intervention. Again, one must conclude that, for whatever

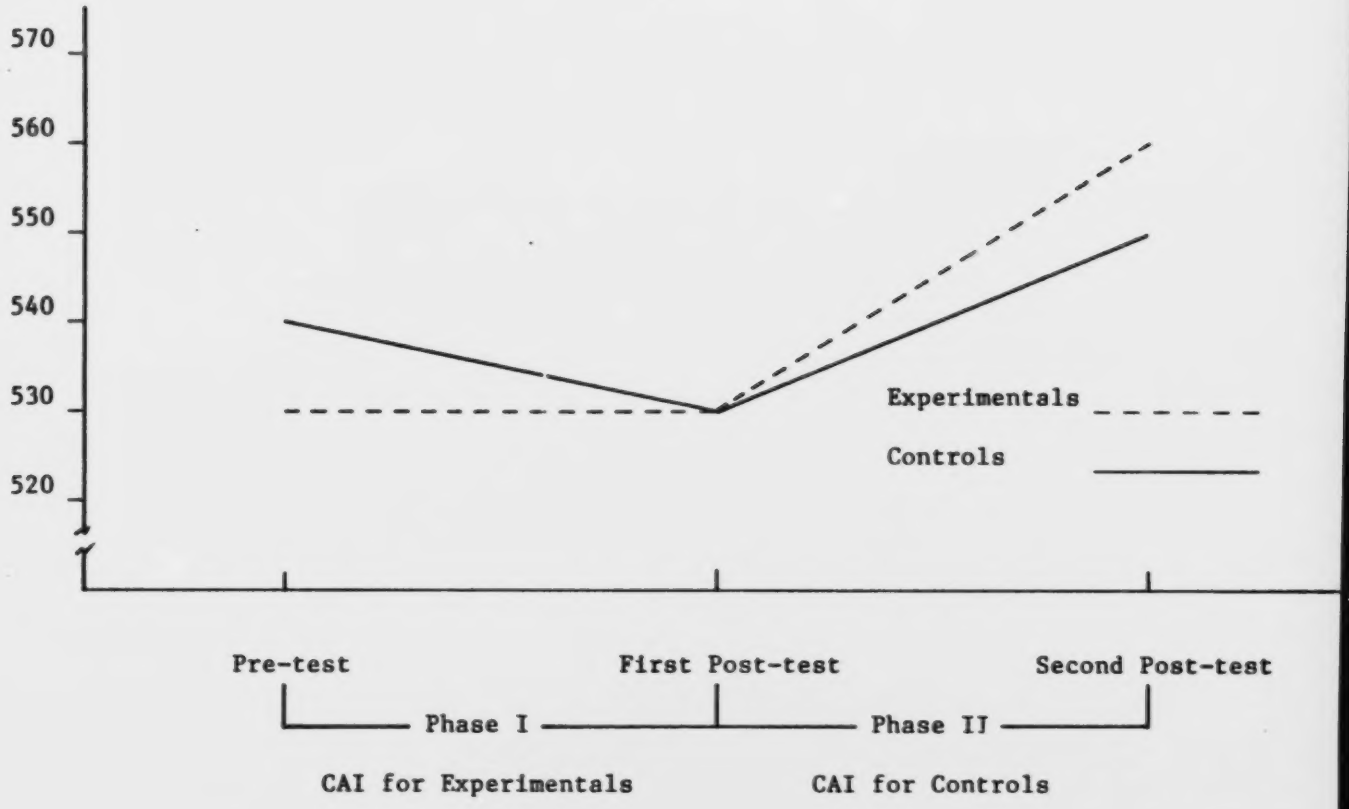

Figure 2. SAT Math Pre- and Post-test Group Mean Score Changes

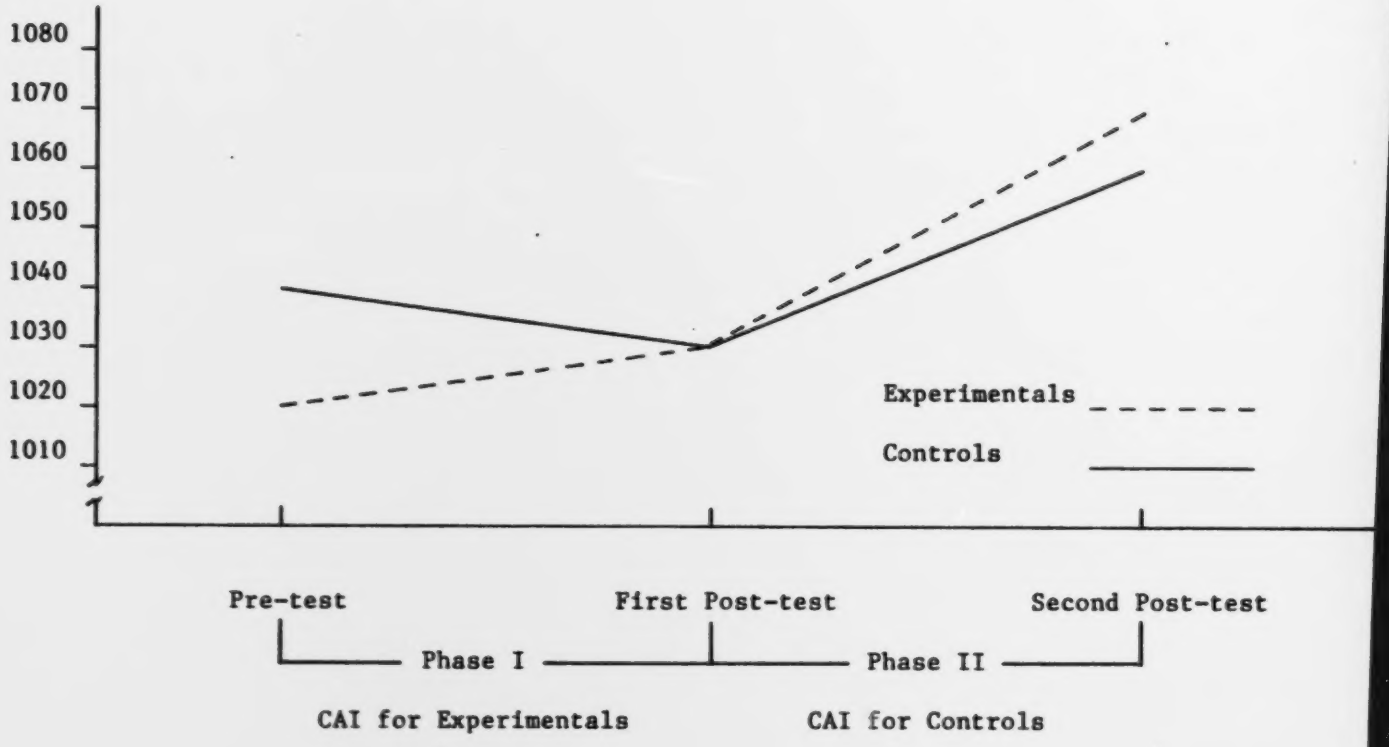

Figure 3. SAT Total Pre- and Post-test Group Mean Score Changes

reasons, members of the entire sample were better able to handle the second post-test than they were the first post-test.

Hypothesis Two

Hypothesis Two states that no difference will be evidenced in subjects' (classified as field-dependent versus field-independent based upon the Witkin GEFT) SAT post-test scores when separated as being above or below the group mean. Analysis of post-test data by means of chi-square revealed no significant difference in student performance following CAI when separated as being field-dependent versus field-independent. This variable bore no relationship beyond chance expectations to SAT scores following CAI intervention. The data upon which this inference is based are reported in Table 6.

Hypothesis Three

Hypothesis Three states that no difference will be evidenced in subjects' (classified as being of high or low socio-economic status relative to the group's median family income) SAT post-test scores as being above or below the group mean. Data reported in Table 7 support this statement. In particular, application of the Mann-Whitney U Test to SAT post-test scores for all students following CAI evidenced no score differences beyond chance expectations. Students from families above the group's median income level did not perform beyond those from families of lower income.

Table 6

Chi-square Applied to Differences Among SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated as Subjects Who Are Field-dependent (F-D) Versus Field-independent (F-I) Based Upon Their Scores From the GEFT

Group	n	GEFT	SAT	χ^2	|p|
1-Experimental	6	F-D	Verbal	3.666	>.05
	16	F-I			
	6	F-D	Math	0.082	>.05
	16	F-I			
2-Control	6	F-D	Verbal	0.757	>.05
	15	F-I			
	6	F-D	Math	0.342	>.05
	15	F-I			
1 & 2 Combined	12	F-D	Verbal	0.559	>.05
	31	F-I			
	12	F-D	Math	0.375	>.05
	31	F-I			

Table 7

Mann-Whitney U Applied to SAT Post-test Scores Reported as Gain or Loss Relative to Pre-test Scores When the Group Is Separated for Family Income Above or Below the Group Median

Group	SAT	Above-R	Below-R	U	z	|p|
1-Experi. n=21	Verbal	101.0	130.0	46.0	---	>.05
	Math	100.0	131.0	45.0	---	>.05
	Total	96.0	135.0	41.0	---	>.05
2-Control n=20	Verbal	76.5	133.5	31.5	---	>.05
	Math	108.5	101.5	35.5	---	>.05
	Total	88.5	121.5	43.5	---	>.05
1 & 2 Combined n=41	Verbal	346.0	515.0	156.0	-1.39	.0823
	Math	423.0	438.0	185.0	-0.63	.2643
	Total	362.0	499.0	172.0	-0.97	.1660

Legend:

Above-R = the sum of the ranks assigned to the scores of students whose family income is above the group median.

Below-R = the sum of the ranks assigned to the scores of students whose family income is below the group median.

Hypothesis Four

Hypothesis Four states that no gender-related difference will be evidenced in SAT post-test scores. Although this hypothesis was stated in the null form, based on the professional literature it could be expected that males would out-perform females in the Math area and that the reverse would be true in the Verbal area. With this in mind, the SAT pre-test data were examined by means of a Mann-Whitney U Test in order to establish the presence of such differences prior to the introduction of CAI. The results of this analysis are reported in Table 8, an inspection of which reveals no significant gender-related differences in the pre-test scores.

While the SAT pre-test scores revealed no significant differences related to gender, one observation merits noting. Seemingly as an artifact of the group assignment process, the males in Group 1 are slightly stronger than the females; while the females in Group 2 are equally as strong as the males. It will be remembered that the initial group assignment was based upon a ranking of SAT Total pre-test scores. Gender was not considered in the group assignments. Consequently, by chance, the assignments which resulted led to the artifact noted above. This would account for the males in Group 1 approaching statistical significance in the Math area, while those in Group 2 do not. Instead, the females tend to score as well as they, thereby causing any male-dominated, gender-related differential performance in Math to be weakened beyond expectations. Such being evident, it should be stated here that a group assignment procedure

Table 8

Mann-Whitney U Applied to SAT Pre-test Score Differences When the Group Is Separated by Gender

Group	SAT	Gender	n	R	U	z	|p|
1-Experi. n=22	Verbal	Male	10	114.0	59.0	---	>.05
		Female	12	139.0			
	Math	Male	10	135.5	39.5	---	>.05
		Female	12	117.5			
	Total	Male	10	126.5	48.5	---	>.05
		Female	12	126.5			
2-Control n=21	Verbal	Male	9	99.0	54.0	---	>.05
		Female	12	132.0			
	Math	Male	9	100.0	53.0	---	>.05
		Female	12	131.0			
	Total	Male	9	101.0	52.0	---	>.05
		Female	12	130.0			
1 & 2 Combined n=43	Verbal	Male	19	410.0	220.0	-0.20	.4207
		Female	24	536.0			
	Math	Male	19	463.0	183.0	-1.10	.1357
		Female	24	483.0			
	Total	Male	19	439.0	207.0	-0.51	.3050
		Female	24	507.0			

which takes this into consideration would have been utilized had gender-related post-test differences been a matter of primary concern. While such differences were of interest, they were secondary to the assessment of the effectiveness of the software under CAI conditions, irrespective of the student's gender.

Further, an analysis of SAT post-test gender-related differences revealed essentially the same results as were apparent with the pre-test scores. There were no gender-related performance differences beyond chance expectations, as reported in Table 9. That is, application of the Mann-Whitney U Test revealed that these results were not statistically significant when separated by gender.

Examining these data in a different manner reinforced even further the above inferences. Application of chi-square to SAT post-test scores, when separated as "high" and "low" achievers relative to the group mean, revealed no significant gender-related differences. The analysis of these data has been reported in Table 10. Consequently, it may be inferred that gender was not a relevant variable in accounting for SAT pre- or post-test performance differences among members of a select group of college-bound students.

Hypothesis Five

Hypothesis Five states that no I.Q.-related difference will be evidenced in SAT post-test scores. Application of a Mann-Whitney U Test to these data revealed that students in the "low" I.Q. range,

Table 9

Mann-Whitney U Applied to SAT Post-test Score Differences When the Group Is Separated by Gender

Group	SAT	Gender	n	R	U	z	|p|
1-Experi. n=22	Verbal	Male	10	103.5	48.5	---	>.05
		Female	12	149.5			
	Math	Male	10	136.5	38.5	---	>.05
		Female	12	116.5			
	Total	Male	10	127.5	47.5	---	>.05
		Female	12	125.5			
2-Control n=21	Verbal	Male	9	92.0	47.0	---	>.05
		Female	12	139.0			
	Math	Male	9	100.5	52.5	---	>.05
		Female	12	130.5			
	Total	Male	9	92.0	47.0	---	>.05
		Female	12	139.0			
1 & 2 Combined n=43	Verbal	Male	19	375.5	185.5	-1.04	.1492
		Female	24	570.5			
	Math	Male	19	481.5	164.5	-1.55	.0606
		Female	24	464.5			
	Total	Male	19	423.5	222.5	-0.13	.4483
		Female	24	522.5			

relative to the group mean, recorded greater gains following CAI than did students in the "high" I.Q. range. Such gains were not statistically significant, however, as shown in Table 11.

Hypothesis Six

Hypothesis Six states that no G.P.A.-related difference will be evidenced in SAT post-test scores. Application of a Mann-Whitney U Test to these data afforded two findings: 1) SAT Verbal post-test score gains are significant beyond the .05 level of confidence for students having G.P.A.s below the group mean, and 2) SAT Math post-test score gains were not sufficiently large to exceed expectations based upon chance occurrences. These data are reported in Table 12.

Hypothesis Seven

Hypothesis Seven states that no difference will be evidenced in SAT post-test scores when separated relative to educational attainment of the parents. Chi-square was applied to these data, and the results are reported in Table 13. No statistically significant relationship between educational attainment of the parents and SAT post-test scores of the progeny was found.

Table 10

Chi-square Applied to Gender Differences Among SAT Post-test Scores When Separated Above and Below the Group Mean

Group	n	Gender	SAT	χ^2	|p|
1-Experi.	10 12	Male Female	Verbal	0.734	>.05
	10 12	Male Female	Math	1.664	>.05
2-Control	9 12	Male Female	Verbal	0.382	>.05
	9 12	Male Female	Math	0.643	>.05
1 & 2 Combined	19 24	Male Female	Verbal	1.091	>.05
	19 24	Male Female	Math	1.996	>.05

Table 11

Mann-Whitney U Applied to SAT Post-test Scores Reported as Gain or Loss Relative to Pre-test Scores When the Group Is Separated for I.Q.s Above or Below the Group Mean

Group	SAT	Above-R	Below-R	U	z	|p|
1-Experi. n=22	Verbal	117.5	135.5	57.5	---	>.05
	Math	108.5	144.5	53.5	---	>.05
	Total	117.5	135.5	57.5	---	>.05
2-Control n=21	Verbal	97.5	133.5	52.5	---	>.05
	Math	113.0	118.0	40.0	---	>.05
	Total	109.5	121.5	43.5	---	>.05
1 & 2 Combined n=43	Verbal	422.5	523.5	223.5	-0.11	.4562
	Math	436.5	509.5	209.5	-0.45	.3264
	Total	443.5	502.5	202.5	-0.62	.2676

Table 12

Mann-Whitney U Applied to SAT Post-test Scores Reported as Gain or Loss Relative to Pre-test Scores When the Group Is Separated for G.P.A.s Above or Below the Group Mean

Group	SAT	Above-R	Below-R	U	z	|p|
1-Experi. n=22	Verbal	99.5	153.5	21.5	---	<.05*
	Math	149.5	103.5	48.5	---	>.05
	Total	108.0	145.0	30.0	---	<.05*
2-Control n=21	Verbal	104.0	127.0	26.0	---	<.05*
	Math	116.5	114.5	38.5	---	>.05
	Total	104.0	127.0	26.0	---	<.05*
1 & 2 Combined n=43	Verbal	437.5	508.5	137.5	-2.21	.0136*
	Math	514.5	431.5	214.5	-0.33	.3707
	Total	425.0	521.0	125.0	-2.52	.0059*

*Reject H_o: Post-test score gains are significant beyond the .05 level of confidence for students having G.P.A.s below the group mean.

Table 13

Chi-square Applied to SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated Relative to Educational Level of the Parents

Group	SAT	Parents	χ^2	$\|p\|$
1-Experimental n=22	Verbal	Fathers	2.556	>.05
		Mothers	0.916	>.05
	Math	Fathers	2.707	>.05
		Mothers	2.674	>.05
	Total	Fathers	1.577	>.05
		Mothers	0.082	>.05
2-Control n=21	Verbal	Fathers	0.542	>.05
		Mothers	0.070	>.05
	Math	Fathers	0.915	>.05
		Mothers	0.643	>.05
	Total	Fathers	1.567	>.05
		Mothers	2.253	>.05
1 & 2 Combined n=43	Verbal	Fathers	0.440	>.05
		Mothers	0.201	>.05
	Math	Fathers	0.686	>.05
		Mothers	0.201	>.05
	Total	Fathers	2.498	>.05
		Mothers	2.168	>.05

Hypothesis Eight

Hypothesis Eight states that no difference will be evidenced in SAT post-test scores when separated relative to occupational status of the parents. A chi-square analysis of SAT post-test scores revealed that the differences between "high" and "low" achievers, as separated above, are not statistically significant. Hence, the null hypothesis was accepted as stated; and the supporting data are reported in Table 14.

Hypothesis Nine

Hypothesis Nine states that no difference will be evidenced in SAT post-test scores when separated relative to students' number of siblings. A chi-square analysis was applied to these data, and the results are reported in Table 15. No statistically significant differences were observed. Number of siblings in the family had no impact beyond chance expectations in generating differential levels of achievement as reflected in students' SAT post-test scores.

Hypothesis Ten

Hypothesis Ten states that no difference will be evidenced in SAT post-test scores when separated relative to each student's proposed college major. Stated as an alternative hypothesis, students choosing college majors in the "hard" sciences would score significantly higher on SAT post-tests than would those selecting career

Table 14

Chi-square Applied to SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated Relative to Occupation of the Parents

Group	SAT	Parents	χ^2	$\|p\|$
1-Experimental n=22	Verbal	Fathers	3.742	>.05
		Mothers	0.180	>.05
	Math	Fathers	0.277	>.05
		Mothers	0.734	>.05
	Total	Fathers	0.277	>.05
		Mothers	0.734	>.05
2-Control n=21	Verbal	Fathers	2.855	>.05
		Mothers	0.076	>.05
	Math	Fathers	4.167	>.05
		Mothers	0.000	>.05
	Total	Fathers	2.303	>.05
		Mothers	2.484	>.05
1 & 2 Combined n=43	Verbal	Fathers	4.625	>.05
		Mothers	0.015	>.05
	Math	Fathers	2.338	>.05
		Mothers	0.245	>.05
	Total	Fathers	1.332	>.05
		Mothers	0.139	>.05

fields outside of these areas. This alternative hypothesis was not sustained, however, in that a chi-square analysis of these data, as reported in Table 16, revealed no statistically significant differences among these sets of scores. Consequently, the null form of the hypothesis was accepted.

Individual Score Changes

A review of the data, particularly as they relate to Hypothesis One, reveals the impact of individual students' score changes following CAI. Although such changes are not reported in the professional literature (which deals specifically with group score changes), there are definite gains and losses observed in these data. These score changes are portrayed rather clearly in Figures 4 through 7.

Figure 4 reveals individual students' gains or losses on either SAT Verbal or Math when pre- and post-test scores are compared following CAI. There is a clear division among these scores at sixty points gained or lost. In the lower group, for instance, the mean number of students evidencing a gain or loss of fifty points or less is twelve; while in the upper group the mean number of students evidencing a gain or loss of sixty points or more is two. It would probably be commonly agreed that a gain or loss of fifty points or less would have little practical consequence for most students; yet one of sixty points or better could have a considerable impact. The relevance of this is treated further under the concept of the standard error of measurement and in the Discussion section of Chapter V.

Table 15

Chi-square Applied to SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated Relative to Subjects' Number of Siblings

Group	SAT	χ^2	$\|p\|$
1-Experimental n=22	Verbal	0.266	>.05
	Math	1.769	>.05
	Total	1.769	>.05
2-Control n=21	Verbal	0.070	>.05
	Math	2.865	>.05
	Total	0.382	>.05
1 & 2 Combined n=43	Verbal	0.277	>.05
	Math	0.277	>.05
	Total	0.115	>.05

Table 16

Chi-square Applied to SAT Post-test Scores Above and Below the Group Mean When the Group Is Separated Relative to Subjects' Proposed College Major

Group	SAT	χ^2	$\|p\|$
1-Experimental n=22	Verbal	0.476	>.05
	Math	1.840	>.05
	Total	2.745	>.05
2-Control n=21	Verbal	0.154	>.05
	Math	1.866	>.05
	Total	0.428	>.05
1 & 2 Combined n=43	Verbal	0.572	>.05
	Math	2.880	>.05
	Total	2.019	>.05

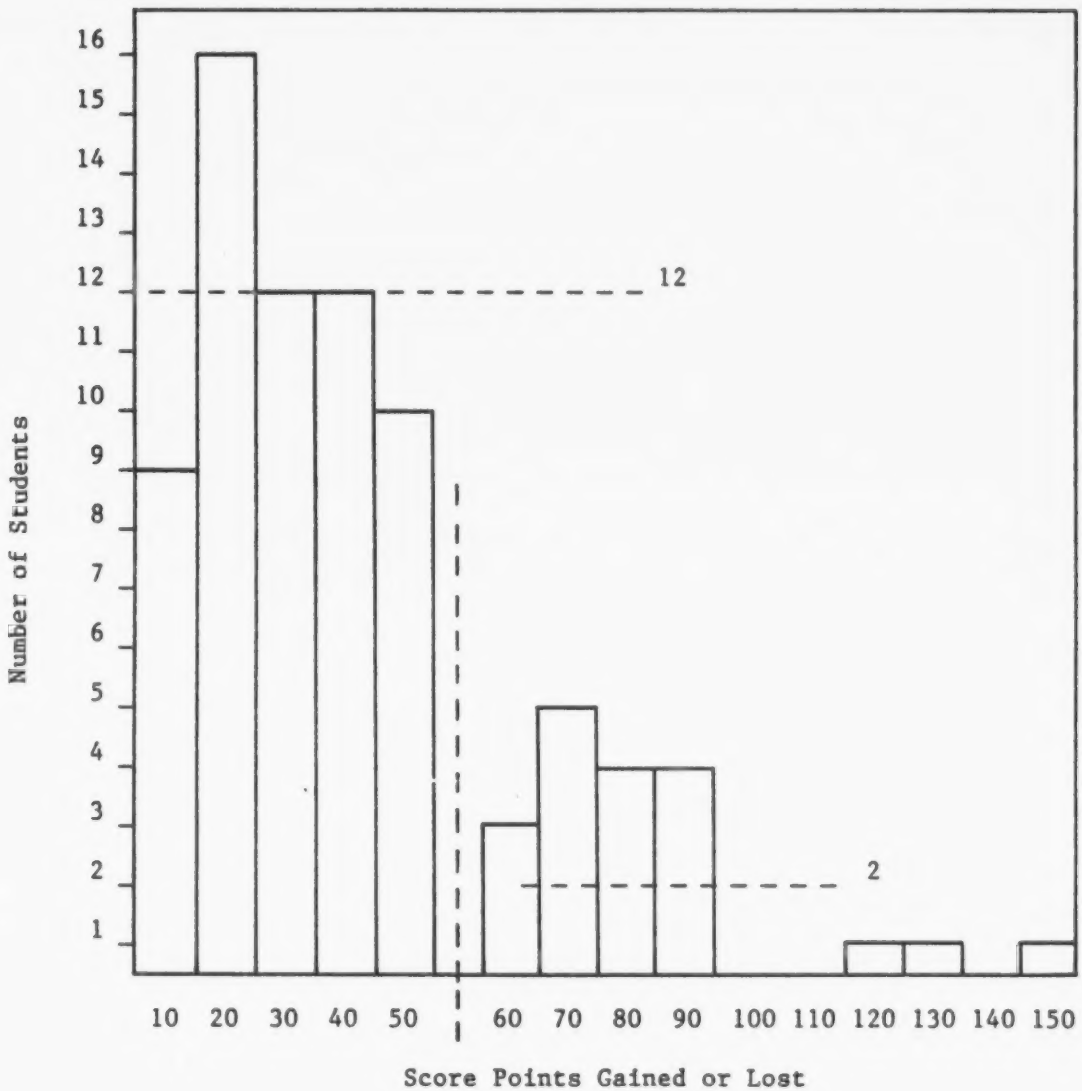

Figure 4. Number of Groups 1 and 2 Students Evidencing a Gain or Loss on Either SAT Verbal or Math When Pre- and Post-test Scores Are Compared Following CAI

As has been noted above, a cut-score of sixty points or more is clearly established in these data. In this regard, fourteen out of forty-three students have been affected; and this means that approximately one-third of the sample evidenced definitive SAT score gains or losses following CAI. Such score instability was not anticipated (from the literature), though it does impact upon the overall relevance of this study. This issue is discussed further in Chapter V.

Presenting examples of the above-noted score instability, Figures 5, 6, and 7 portray graphically the more extreme score changes for three individual students. Rather dramatic post-CAI score gains and losses are evident. Such score losses are not reported in the publisher's literature; while on the other hand, these are the kinds of score gains which are reported as the basis for claims of product effectiveness. Hence, this raises the question of which students might benefit most from the use of these products. In this respect, consideration was given to the possibility of developing an individual student profile from these data. While this would certainly seem desirable, such was not feasible in the present study due to the limited number of students evidencing the more extreme score changes.

In order for these individual score changes to be put in proper perspective and assessed as to their more realistic meaning, they must be considered relative to the concept of the standard error of measurement. These data were so analyzed; and the results are reported in Table 17, wherein these differences in the standard error of measurement are reflected in sets of SAT scores for students with

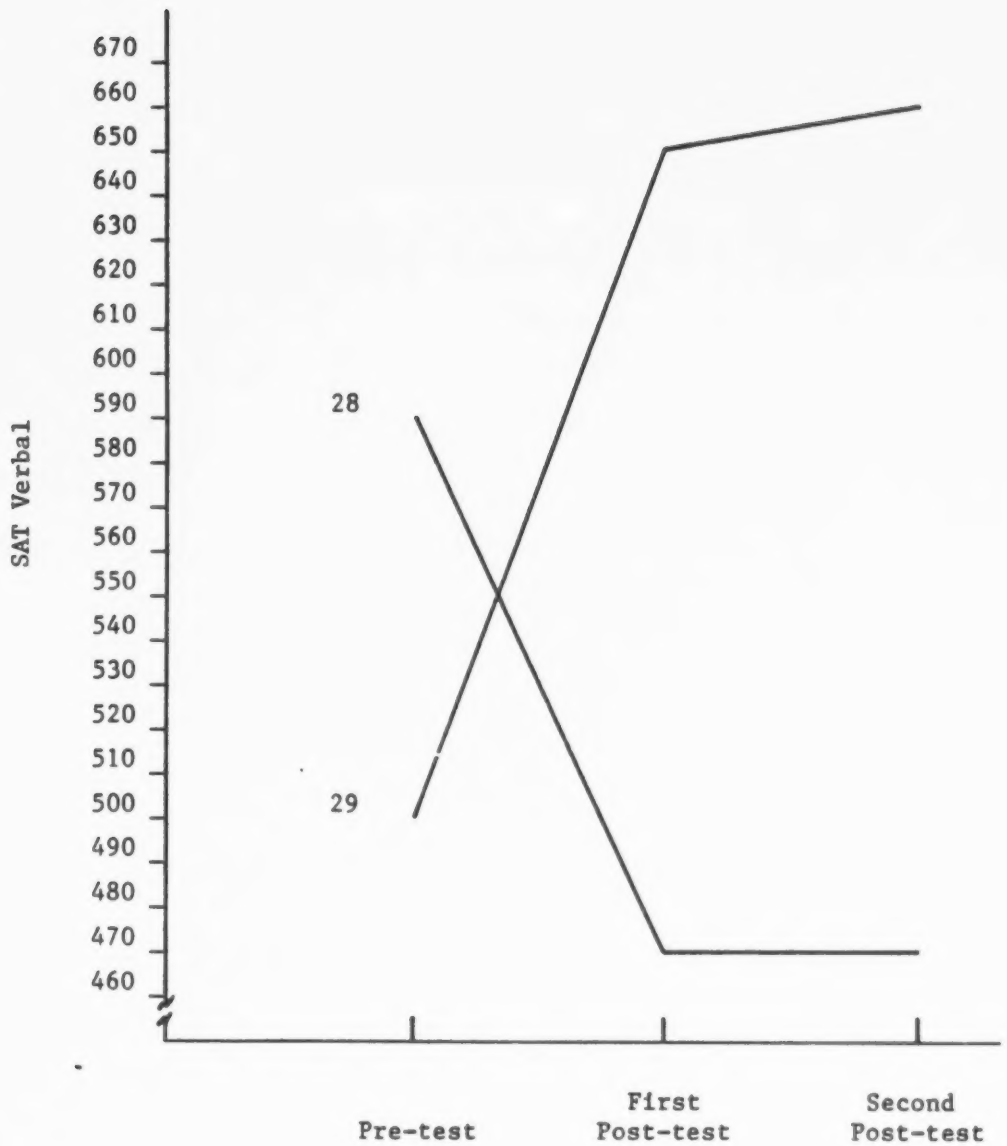

Figure 5. Group 1 Students 28 and 29 Evidencing Extreme SAT Verbal Post-test Score Changes Following CAI

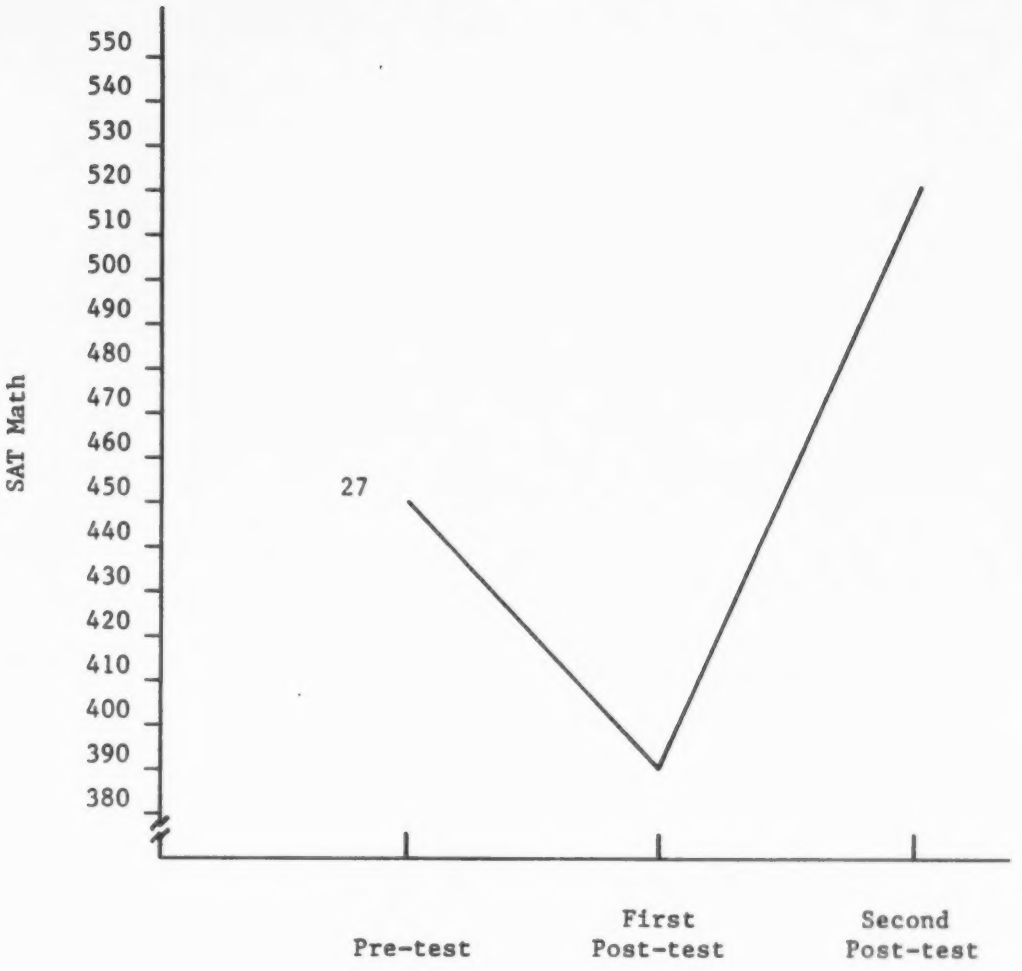

Figure 6. Group 2 Student 27 Evidencing Extreme SAT Math Post-test Score Change Following CAI

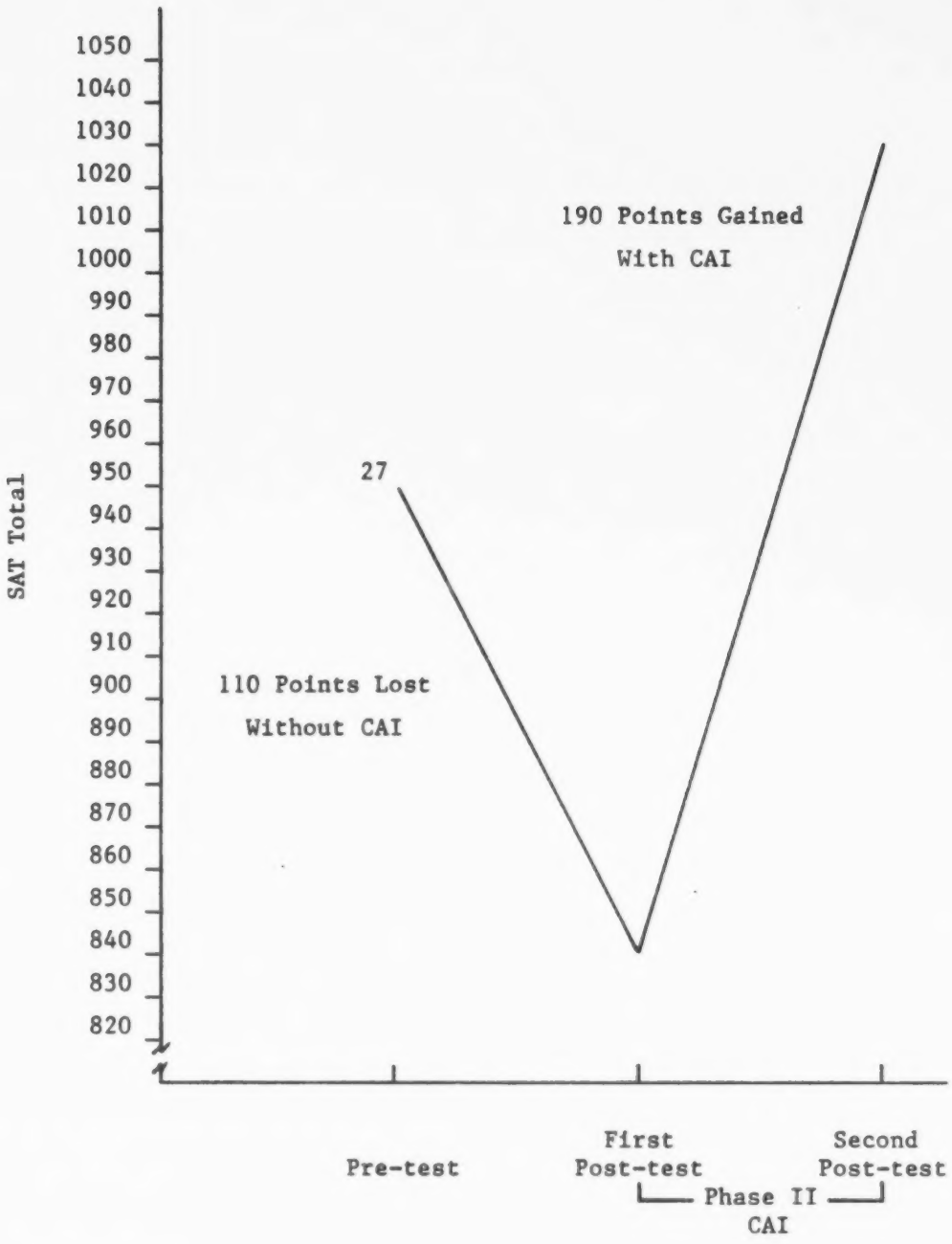

Figure 7. Group 2 Student 27 Evidencing Extreme SAT Total Post-test Score Change Following CAI

and without CAI. Inspection of these data reveals no clear pattern as to whether the scores vary more with or without CAI. It should be noted, however, that in 5 out of 9 comparisons the difference is zero. In addition, application of Spearman Rho to these differences yielded a value of .64, which is significant beyond the .05 level of confidence. Thus it would appear from this analysis that CAI involving Krell SAT software has little impact relative to meaningful score change.

Retention

Looking specifically at retention, it would seem important in any study of learning, whatever the methodology, to obtain a measure of how well students retain what they have learned. Phase II of the design, wherein Group 1 students took both SAT post-tests, provided such a measure. As is apparent in the results reported in Table 18, which reflects a Mann-Whitney U Test application to these data, retention scores continued to rise following the completion of CAI. While the most notable improvement is reflected in the Math retention scores (as previously detailed in Figure 2, page 48), even these gains are not beyond chance expectations. Hence, it would have to be concluded from an overall inspection of these retention measures that the increases noted are not statistically significant.

Table 17

Standard Error of Measurement Derived From SAT Pre- and Post-test Score Differences Following CAI

Group	SAT	Std. Error With CAI	Std. Error Without CAI	Difference
1-Experi. n=22	Verbal	60	40	20
	Math	50	50	0
	Total	60	60	0
2-Control n=21	Verbal	50	50	0
	Math	50	60	10
	Total	70	80	10
1 & 2 Combined n=43	Verbal	50	40	10
	Math	50	50	0
	Total	70	70	0

Table 18

Mann-Whitney U Applied to Group 1 SAT Post-test Retention Scores One Month Following Completion of the CAI

Group	SAT	First-R	Second-R	U	z	|p|
1-Experi. n=22	Verbal	504.0	486.0	233.0	-.21	.4168
	Math	454.0	536.0	201.0	-.96	.1685
	Total	475.0	515.0	222.0	-.47	.3192

Legend:

First-R = the sum of the ranks assigned to the students' SAT scores from the first post-test immediately following the CAI.

Second-R = the sum of the ranks assigned to the students' SAT scores from the second post-test one month following the CAI.

Correlational Relationships

In addition to the analyses thus far reported, Spearman Rho was applied to SAT post-test scores and several other measures (viz., O-LSAT I.Q.s, G.P.A.s, and family income) in order to determine the degree of correlational relationship. The intent was to assess the predictive value of these measures relative to SAT post-test performance. An inspection of the data shown in Table 19 reveals that I.Q. and G.P.A. were the strongest predictors of success on the SAT, especially on the Math subtest. In particular, high scores on the SAT post-tests were positively and significantly related to high I.Q. and G.P.A., with both being beyond the .05 and .01 levels of confidence. Family income, however, tended to bear a weak, inverse relationship to SAT post-test scores.

Summary

This chapter presented the data analyses and report of the findings. In summary, nine out of ten of the hypotheses tested were accepted in the null form. Only a portion of Hypothesis Six was rejected; that is, students with G.P.A.s below the group mean achieved higher post-CAI Verbal scores than did those with G.P.A.s above the mean. Beyond this, the only other relationships found to be significant were: SAT post-test scores in all areas when correlated with I.Q., and in Math and Total score when correlated with G.P.A.

Table 19

Spearman Rho Applied to SAT Post-test Scores and O-LSAT (I.Q.), G.P.A., and Family Income (F.I.) to Determine the Degree of Correlational Relationship

Group	SAT	O-LSAT	G.P.A.	F.I.
1-Experimental n=22	Verbal	.660**	.100	-.030
	Math	.584**	.641**	-.069
	Total	.807**	.483*	-.040
2-Control n=21	Verbal	.782**	.201	.197
	Math	.567**	.393*	.017
	Total	.765**	.399*	.112
1 & 2 Combined n=43	Verbal	.711**	.114	.067
	Math	.607**	.508**	-.001
	Total	.799**	.348*	.059

**Reject H_o: Correlational relationship is significant beyond the .01 level of confidence.

*Reject H_o: Correlational relationship is significant beyond the .05 level of confidence.

CHAPTER V

SUMMARY, FINDINGS, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Summary

The purpose of this study was to assess empirically the most expensive SAT microcomputer software being marketed at the time this research was initiated, and about which extensive claims of success have been made by the publisher. This new technology challenges the professional literature of forty years which reported that no SAT coaching methodology has yet appeared which could produce statistically significant results of practical consequence. Even the best programs available have produced only modest gains, and these have seemed not to justify the time and expense involved. Consequently, it was deemed important to test the effectiveness of CAI (utilizing selected software) in raising SAT scores under controlled conditions involving two matched groups of college-bound juniors from the Escambia County (Florida) School District.

The conceptual framework for this study evolved principally from Skinner's research in operant conditioning. This model forms the basis for much of the SAT software currently being developed. There is stimulus presentation, student response, and immediate feedback associated with reinforcement. The desired operant would be to teach the student, through programmed CAI, to select the correct choice on the

SAT. In addition to Skinner's work, Witkin's research in cognitive styles (i.e., field-dependence versus field-independence) provided a secondary conceptual model. The intent here was to relate learning styles to SAT outcome measures as a consequence of the CAI. Lastly, certain assumptions and limitations were stipulated; and specific hypotheses were formulated and tested through appropriate statistical procedures.

A review of the related literature encompassed seven major topics. These were as follows: 1) historical use of college admission scores, 2) effects of tutoring upon SAT scores, 3) attacks upon the SAT, 4) defense of the SAT, 5) potential impact of the microcomputer, 6) computer-assisted instruction, and 7) cognitive styles. As will be noted, four of these areas deal specifically with the SAT; while the others relate to use of the microcomputer and its potential impact upon individual students' learning styles.

The methodology for this study incorporated a two-phased, classical pre- and post-test design. Phase I involved a comparison of baseline and outcome measures obtained from pre- and post-testing of students systematically assigned to Experimental and Control groups. Phase II permitted a replication of the CAI impact assessment in Phase I, plus the acquisition of retention measures following CAI for the Experimentals. Data collection consisted of administration of one SAT pre-test, two SAT post-tests, the Otis-Lennon School Ability Test (I.Q.), the Witkin Group Embedded Figures Test (cognitive styles), and the use of a student questionnaire to obtain demographic

information. The latter involved the student's gender, his/her G.P.A. in the district's core curriculum, the educational level and occupational status of the parents, family income, number of siblings, and the student's proposed college major. Data analysis utilized nonparametric techniques as follows: the Mann-Whitney U Test, chi-square, and Spearman Rho. All SAT coaching sessions were conducted in a lab setting involving Krell software and Apple IIe microcomputers. The "coaching" or instructional method was limited strictly to CAI.

Findings

The findings of this study may be listed as follows:

1. Experimental subjects receiving CAI involving Krell software did not significantly out-perform Control subjects as reflected in SAT post-test score gains.

2. SAT post-test scores were not significantly different for groups classified as field-dependent versus field-independent based upon scores from the Witkin GEFT.

3. SAT post-test scores were not significantly different for students whose socio-economic status of the family was classified as high or low relative to the group's median income.

4. Gender-related SAT post-test score differences were not statistically significant.

5. I.Q.-related SAT post-test score differences were not statistically significant.

6. When separated relative to the group's mean G.P.A., "low"

students significantly out-performed "high" students in SAT Verbal post-test score gains. Math score gains following CAI, however, were not significant.

7. No significant difference was found in SAT post-test score gains when students were separated relative to high or low educational attainment of the parents.

8. No significant difference was found in SAT post-test score gains when students were separated relative to high or low occupational status of the parents.

9. SAT post-test scores were not significantly different when students were separated relative to number of siblings.

10. SAT post-test scores were not significantly different when students were separated relative to proposed college major.

Conclusions

The findings of this study indicate essentially two things: 1) that Krell SAT software utilized under controlled CAI conditions produces only modest group score gains, and 2) that individual student score gains under these conditions, by contrast, may be high. That is, group SAT score gains will seldom exceed chance expectations, which means that such gains will be of little practical consequence. This inference only serves to reinforce the long-standing position of the professional literature. On the other hand, individual students may evidence significant gains, which is entirely consistent with the claims embodied in the publisher's literature.

It also merits attention, relative to the comparison of pre- and post-test scores following CAI, that the results in Phase II of this study replicated precisely those recorded in Phase I. The observed gains, as outcome measures, were at no point statistically significant. This says rather clearly, individual results notwithstanding, that the CAI provided by the Krell 1981/82 SAT preparatory series software was not effective in raising group mean post-test scores beyond chance expectations.

Discussion

Other observations from the findings afford several notations. For instance, students having I.Q.s below the sample mean (though they are certainly quite bright) tended to out-perform those in the upper group in terms of relative SAT score gain. This would tend to support the common generalization that students who score in the "high average" or "bright normal" range strive to over-achieve in competition with students of greater ability. Second, students whose G.P.A.s were below the group mean significantly out-performed academically stronger students in terms of relative gains in the Verbal area. This again seems to be related to tendencies toward over-achievement. Third, students whose family income fell below the group median tended to out-perform students from families of higher income as reflected in relative gains in both the Verbal and Math areas, though these gains were not statistically significant. Striving to achieve and grow socially, which is certainly reinforced in

this culture and directly related to college attendance, is readily apparent.

The second observation cited in the above paragraph seems to merit special note. Is this, for instance, a "true" finding; or is it an artifact of the data? As has already been stated elsewhere, it is the only statistically significant finding of the ten empirical hypotheses tested. The present researcher is of the opinion that it is more artifact than fact; and there are at least four factors or conceptual inferences which influence this judgment. First is the concept of the asymptotic limit, which is relevant here in that students whose scores are above the group mean have a narrower range for potential improvement; while those whose scores fall below the mean have a broader range for improvement. Second is the notion of score regression toward the mean. This is particularly relevant in that high scores tend to come down on post-tests; while low scores tend to move up toward the mean. Third is the common expectation that at least one in twenty empirical tests of statistical significance at the .05 level of confidence will result in a positive outcome; that is, the null hypothesis will be rejected. Fourth is the very relevant observation that five out of six of the bilingual students in the present sample were in the low-I.Q./high gain group, meaning that fifteen hours of intensified study (most of which was spent upon Verbal tasks and vocabulary-building) had a positive impact upon improving SAT Verbal post-test scores. As stated earlier, this researcher accepts the inference that these factors were much

more relevant in influencing the reported positive outcome than was the particular instructional methodology under investigation.

As to the relevance of demographic variables studied, it should be specifically stated that observed differences in SAT post-test score gains were not statistically significant when students were separated relative to family income, educational level of the parents, occupational status of the parents, number of siblings, and the student's proposed college major. Though stated in the null form, it had been assumed, since there is precedent for such in the professional literature involving average students, that all of these variables would influence observed differences in post-test score gains following CAI. In no instance, however, were these assumptions sustained. Consequently, hypotheses involving these variables were accepted as stated.

Strong positive relationships were observed between SAT post-test scores following CAI and the variables general intelligence (I.Q.) and G.P.A. in the district's core curriculum. In particular, these relationships were significant beyond either the .05 or the .01 level of confidence. This finding is entirely consistent with notations in the literature, wherein high school G.P.A. is reported to be the number one predictor of first-year success in college, with SAT scores following a close second. General intelligence has long been noted as highly correlated with academic achievement at all levels. High SAT scores are only one more indicator of such ability, motivation, and achievement -- albeit an academic bridge between the

secondary and post-secondary levels.

As already indicated, the effectiveness of the SAT software package produced for use with the microcomputer has not been sustained where meaningful group mean score gains are concerned. For isolated individual score gains, however, the question yet remains open. This obviously raises the issue of practicality, especially where the expenditure of funds is concerned. To obtain a complete setup for an individual student would be prohibitive for most families; and there are no assurances that this particular student would benefit from the training. Beyond this, the expenditure of public funds toward the goal of attaining significant group SAT score gains could not be defended from the findings of this research.

There are other questions which also merit attention: 1) Why are the SAT scores for some students so unstable under repeated testing? 2) How does one account for major score losses following CAI? 3) Is differential test difficulty a matter of legitimate concern? Consideration must also be given to a discussion of the concept of the standard error of measurement and its particular relevance to score changes reported in the findings of this study. This concept is especially important where coaching methodologies are being evaluated. Score changes, irrespective of the instructional methodology being employed, must exceed the standard error of measurement in order to have any practical significance. The minimal gains reported in the coaching literature fall well within the standard error of measurement recorded among the findings of this study. Given this

situation, it would seem meaningless to conclude that a coaching methodology which produced results within the commonly reported ranges had any practical value.

As to the second question posed, major score losses probably have little or nothing to do with the CAI per se. This position would not be difficult to defend. Far more likely would be the impact of the testing situation itself, the test content, and/or the social and emotional status of the student at the time. There is little question that testing situations vary considerably, particularly as related to physical comfort, distractions, and real or imagined stress. Arriving late, for instance, could in itself generate all kinds of problems. More important, however, seems to be the matter of test content. As in any testing experience, the student's performance is greatly determined by familiarity with the specific content being sampled. On the SAT, in particular, the range of item content is extremely broad. Consequently, the typical college-bound student would seem to be vulnerable to score fluctuations under repeated testing. Such was clearly true in this study where all participants took the SAT three times, and approximately one-third of the sample experienced rather dramatic score changes.

There is also the matter of guessing. Some students guess freely; others are fearful of attempting any item where they experience uncertainty. With the SAT, where each item ultimately equals ten points in the scaled score (which is the reported score), guessing could quickly alter one's score. Not only is there a correction

factor which penalizes one for incorrect guessing, there is the converse possibility of raising one's score spuriously through correct guessing. Some students report consciously experimenting with this factor, ultimately attempting to improve their scores. How successful they are is not a matter of record.

Of possibly greater importance, however, is the third question. This question concerns differential test difficulty. Does the SAT change in difficulty from one administration to the next? Have the different forms of the test been statistically equated as to item content difficulty? For a given student, the difficulty level of the test does vary; for the test itself, however, the item content difficulty does not vary. All editions of the SAT are constructed to meet strict specifications for content and difficulty level. The contrary situation would clearly throw into serious question the validity and reliability of the SAT college admissions program. While this could happen, it would not be because the tests are declared sufficiently unequal in difficulty to warrant and sustain a student's claim of discriminatory violation of individual rights. Equality of the tests is defensible; this is not the problem. Moreover, it clearly is not the answer to the question of notable score changes. Whatever the answer, it lies elsewhere.

Recommendation to the Field

Before funding and initiating an SAT coaching program involving a CAI methodology, educators should review carefully the professional

literature relevant to this question. In this same respect, SAT software publishers' claims of post-test score gains by individual students following only a few hours of CAI should be considered with caution. When reviewing SAT software for possible purchase, publishers should be requested to provide the results of controlled validation studies reflecting significant group gains in post-test scores. All reported SAT post-test score gains, irrespective of the coaching methodology employed, should be evaluated relative to the standard error of measurement appropriate to the data. Reports of SAT software effectiveness where students' post-test score losses are either ignored or not cited should be carefully examined. Parents who seek advice relative to the purchase of SAT software for use by their own children should be cautioned that its effectiveness for individual students is uncertain and not yet predictable.

Recommendation for Further Research

Additional controlled studies in this area of research should be conducted. Primary focus should be placed upon the identification of individual students who are likely to benefit most from SAT software utilized under a short-term CAI methodology. Herein emphasis should be directed toward the score gains of these particular students with the intent of developing an identifying or predictive profile. Researchers who engage in these studies are additionally encouraged to employ designs which incorporate replication and retesting under a varied retention schedule.

REFERENCES

Alderman, Donald L., & Powers, Donald E. (1980). The effects of special preparation on SAT-Verbal scores. American Educational Research Journal, 17(2), 239-251.

Allport, Gordon W. (1961). Pattern and growth in personality. New York: Holt, Rinehart & Winston.

Angoff, William H. (1973-74). A reminder for educators: Seven points to remember about the SAT. College Board Review, 90, 5.

An interview with Stanley Kaplan. (1984). Career World, 13(1), 18.

Astin, Alexander W. (1965). Who goes to college? Chicago: Science Research Associates.

Atkinson, R. C., & Wilson, H. A. (1969). Computer-assisted instruction: A book of readings. New York: Academic Press.

Biemiller, Lawrence. (1982). Microcomputers expected to bring radical changes in testing. The Chronicle of Higher Education, 25 (11), 4.

Black, John A. (1979). When is a study not a study? An examination of guidelines for improving SAT scores. Phi Delta Kappan, 60(7), 487-489.

Blaisdell, F. J. (1976-77). Historical development of computer assisted instruction. Journal of Educational Technology Systems,

5, 155-170.

Bond, Lloyd. (1981). Before the test: Coaching, equity, and admissions. New Directions for Testing and Measurement, 9, 53-59.

Boyd, William M. (1980). Black students out-perform their SAT scores: Excellent secondary school preparation can impart important knowledge and attitudes. Independent School, 39(4), 35-37.

Bracey, Gerald W. (1980). The SAT, college admissions, and the concept of talent: Unexplained myths, unexplained perceptions, and needed explorations. Phi Delta Kappan, 62(3), 197-199.

Bunderson, C. V. (1972). Mainline CAI, necessary, but not oppressive. Proceedings of the American Federation of Information Processing Societies Conference, 399-405.

Burns, Richard W. (1976). Minorities, instructional objectives and the SAT. Educational Technology, 16(6), 39-41.

Coffman, William E. (1980). The Scholastic Aptitude Test: An historical perspective. College Board Review, 117, 8-11.

College Entrance Examination Board. (1979). Board reaffirms its position that "coaching" for SAT is not likely to improve students' scores. College Board News, 34(9), 6.

College Entrance Examination Board. (1981). Board reaffirms its position that "coaching" for SAT is not likely to improve students' scores. College Board News, 36(3), 9-10.

Coughlin, Ellen K. (1982). SAT's test: (a) aptitude (b) achievement (c) a and b. The Chronicle of Higher Education, 24(19), 18.

Cross, K. Patricia. (1976). Accent on learning. San Francisco: Jossey-Bass Publishers.

Cullen, Tom. (1982, October). SAT tutoring programs: Give them an incomplete. Personal Computing, pp. 121-129.

Dalton, Starrette. (1974). Predictive validity of high school rank and SAT scores for minority students. Educational and Psychological Measurement, 34(2), 367-370.

Donlon, Thomas F. (1981). The SAT in a diverse society: Fairness and sensitivity. College Board Review, 122, 16-21, 30-32.

Dwyer, T. A. (1972). Teacher/student authored CAI using the NEW-BASIC system. Communications ACM, 15, 21-28.

Edmonds, Ronald R. (1982). Comments on "Should we relabel the SAT... or replace it?" New Directions for Testing and Measurement, 13, 51-57.

Elliott, Bob. (1978). Two changes for ACT assessment. Activity, 16(3), 6.

Farr, Roger, & Olshavsky, Jill Edwards. (1980). Is minimum competency testing the appropriate solution to the SAT decline? Phi Delta Kappan, 61(8), 528-530.

Federal Trade Commission. (1981). Staff report on the Federal Trade Commission investigation of coaching for standardized admission tests. Boston, Massachusetts: U.S. Government Printing Office.

Federal Trade Commission. (1978). The effects of coaching on standardized admission examinations: Staff memorandum of the Boston regional office of the Federal Trade Commission. Boston, Massachusetts: U.S. Government Printing Office.

Gallagher, Elizabeth. (1980). Preparing for the S.A.T.: Applications of a research project. Clearing House, 53(5), 230-232.

Gifford, Charles S., & Fluitt, John L. (1980). How to make your students testwise: Practice, practice, practice. American School Board Journal, 167(10), 29, 40.

Guilford, J. P. (1956). Fundamental statistics in psychology and education (3rd ed.). New York: McGraw-Hill Book Company, Inc.

Harris, William U. (1976). The SAT score decline: Facts, figures and emotions. Educational Technology, 16(6), 15-20.

Hillje, Barbara Brown. (1980). Some horsesense about raising SAT scores. English Journal, 69(6), 29-31.

Holloway, Charles M. (1971). College entrance examination board. In Lee C. Deighton (Ed.), The encyclopedia of education: Vol. 2. New York: The Macmillan Company & The Free Press.

Jackson, Rex. (1980). The Scholastic Aptitude Test: A response to

Slack and Porter's "Critical appraisal." Harvard Educational Review, 50(3), 382-391.

Jencks, Christopher, & Crouse, James. (1982). Aptitude vs. achievement: Should we replace the SAT? Public Interest, 67, 21-35.

Jencks, Christopher, & Crouse, James. (1982). Should we relabel the SAT... or replace it? New Directions for Testing and Measurement, 13, 33-49.

Karney, Wesley. (1980). The SAT: A parent speaks. College Board Review, 117, 18-20.

Kinder, Dennis Ray. (1979). "On further examination" of the SAT report. Contemporary Education, 50(3), 172-175.

Kingston, Robert J. (1979, September). Board reaffirms its position that "coaching" for SAT is not likely to improve students' scores. The College Board News, p. 6.

Kogan, N. (1971). Educational implications of cognitive styles. In G. S. Lesser (Ed.), Psychology and educational practice. Glenview, Illinois: Scott-Foresman.

Krell Software Corporation. (1981). Krell college board SAT exam preparation series. Stoney Brook, New York: Krell and Associates, Inc. (Pamphlet)

Kubistant, Tom. (1981). Test performance: The neglected skill. Education, 102(1), 53-55.

Lockhart, Kathleen A. (1981). Computer-managed instruction in the Navy: IV. The effects of test item format on learning and knowledge retention (Report No. NPRDC-TR-81-8). San Diego, CA: Navy Personnel Research and Development Center.

Magidson, E. M. (1977). One more time: CAI is not dehumanizing. Audiovisual Instruction, 22, 20-21.

Messick, Samuel. (1981). The controversy over coaching: Issues of effectiveness and equity. New Directions for Testing and Measurement, 11, 21-53.

Messick, Samuel. (1970). The criterion problem in the evaluation of instruction: Assessing possible, not just probable, intended outcomes. In M. C. Wittrock & D. E. Wiley (Eds.), The evaluation of instruction: Issues and problems. New York: Holt.

Messick, Samuel, & Jungeblut, Ann. (1981). Time and method in coaching for the SAT. Psychological Bulletin, 89(2), 191-216.

Moulthroup, Robert. (1981). Legislative proposals to regulate standardized admissions testing: Political issues, educational truths, and the political process. Peabody Journal of Education, 58(2), 72-83.

Nairn, Allan. (1980). The reign of ETS: A brief summary of the Ralph Nader report on the Educational Testing Service. Today's Education, 69(2), 58-64.

Neil, Marion. (1975). Cognitive style: A new aspect to instructional technology. In Arthur M. Cohen & Florence B. Brawer (Eds.), New directions for community colleges (No. 9). San Francisco: Jossey-Bass Publishers.

Nievergelt, J. (1975). Interactive systems for education: The new look at CAI. In O. Lecarme & R. Lewis (Eds.), Computers in education. Amsterdam: North-Holland Publishing Co.

Oltman, P. K., Raskin, E., & Witkin, H. A. (1971). Group embedded figures test. Palo Alto, California: Consulting Psychologists Press, Inc.

Powers, Donald E. (1981). Students' uses of and reactions to alternative methods of preparing for the SAT. Measurement and Evaluation in Guidance, 14(3), 118-126.

Rovezzi-Carroll, Susan, & Thompson, Donald L. (1980). Forecasting college success for low-income students. Journal of College Student Personnel, 21(4), 340-343.

Ryor, John. (1977). Declining SAT scores. Today's Education, 66, 4, 6, 8.

Savage, David G. (1978). The long decline in SAT scores. Educational Leadership, 35(4), 290-293.

Siegel, Sidney. (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Book Company, Inc.

Skinner, B. F. (1953). Science and human behavior. New York: The Macmillan Company.

Skinner, B. F. (1954). The science of learning and the art of teaching. Harvard Educational Review, 24, 86-97.

Slack, Warner V., & Porter, Douglas. (1980). The Scholastic Aptitude Test: A critical appraisal. Harvard Educational Review, 50 (2), 154-175.

Slack, Warner V., & Porter, Douglas. (1980). Training, validity, and the issue of aptitude: A reply to Jackson. Harvard Educational Review, 50(3), 392-401.

Suppes, P., & Macken, E. (1978). The historical path from research and development to operational use of CAI. Educational Technology, 18, 9-11.

Thompson, Scott D. (1979). When a criticism is not a criticism. Phi Delta Kappan, 60(7), 490-491.

Thompson, Scott, & DeLeonibus, Nancy. (1978). Guidelines for improving SAT scores. Reston, Virginia: National Association of Secondary School Principals.

Venable, T. C. (1981). Declining SAT scores: Some unpopular hypotheses. Phi Delta Kappan, 62(6), 443-445.

Will coaching help your SAT score? (1984). Career World, 13(1), 17-18.

Winter, David G. (1977). Motivational factors in the SAT score decline. New York, New York: College Entrance Examination Board.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1974). Psychological differentiation. Potomac, Maryland: Erlbaum.

Witkin, H. A., & Moore, C. A. (1974, April). Cognitive style and the teaching-learning process. Paper presented at annual meeting of American Educational Research Association, Chicago, Illinois.

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1975). Field-dependent and field-independent cognitive styles and their educational implications. Princeton, New Jersey: Educational Testing Service.

Witkin, H.A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). A manual for the embedded figures test. Palo Alto, California: Consulting Psychologists Press, Inc.